National Library of Canada
Bibliothèque nationale du Canada

Canadian Theses Service
Services des thèses canadiennes

Ottawa, Canada
K1A 0N4

CANADIAN THESES
THÈSES CANADIENNES

NOTICE

The quality of this microfiche is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us an inferior photocopy.

Previously copyrighted materials (journal articles, published tests, etc.) are not filmed.

Reproduction in full or in part of this film is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30. Please read the authorization forms which accompany this thesis.

AVIS

La qualité de cette microfiche dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de qualité inférieure.

Les documents qui font déjà l'objet d'un droit d'auteur (articles de revue, examens publiés, etc.) ne sont pas microfilmés.

La reproduction, même partielle, de ce microfilm est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30. Veuillez prendre connaissance des formules d'autorisation qui accompagnent cette thèse.

THIS DISSERTATION
HAS BEEN MICROFILMED
EXACTLY AS RECEIVED

LA THÈSE A ÉTÉ
MICROFILMÉE TELLE QUE
NOUS L'AVONS REÇUE

NL 339 (r. 86/01)

Canada

336

0-315-27067-5

National Library of Canada
Bibliothèque nationale du Canada

Canadian Theses Division
Division des thèses canadiennes

Ottawa, Canada
K1A 0N4

## PERMISSION TO MICROFILM — AUTORISATION DE MICROFILMER

• Please print or type — Écrire en lettres moulées ou dactylographier

Full Name of Author — Nom complet de l'auteur

Elizabeth Trent Nadon

| Date of Birth — Date de naissance | Country of Birth — Lieu de naissance |
|---|---|
| March 20, 1962 | Canada |

Permanent Address — Résidence fixe

18 Bluebird Crescent
Sherwood Park, Alberta
T8A 0E7

Title of Thesis — Titre de la thèse

Early Indices of Grade One Academic ~~Achievement~~ Performance

University — Université

University of Alberta

Degree for which thesis was presented — Grade pour lequel cette thèse fut présentée

M.Ed.

| Year this degree conferred — Année d'obtention de ce grade | Name of Supervisor — Nom du directeur de thèse |
|---|---|
| 1983 | Dr. H. Janzen |

Permission is hereby granted to the NATIONAL LIBRARY OF CANADA to microfilm this thesis and to lend or sell copies of the film.

The author reserves other publication rights, and neither the thesis nor extensive extracts from it may be printed or otherwise reproduced without the author's written permission.

L'autorisation est, par la présente, accordée à la BIBLIOTHÈQUE NATIONALE DU CANADA de microfilmer cette thèse et de prêter ou de vendre des exemplaires du film.

L'auteur se réserve les autres droits de publication; ni la thèse ni de longs extraits de celle-ci ne doivent être imprimés ou autrement reproduits sans l'autorisation écrite de l'auteur.

| Date | Signature |
|---|---|
| June 23/83 | Elizabeth Nadon |

NL-91 (4/77)

THE UNIVERSITY OF ALBERTA

# EARLY INDICES OF GRADE ONE ACADEMIC PERFORMANCE

by

(C) Elizabeth T. Nadon

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES AND RESEARCH

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF EDUCATION

IN

EDUCATIONAL PSYCHOLOGY

DEPARTMENT OF EDUCATIONAL PSYCHOLOGY

EDMONTON, ALBERTA

FALL, 1983

THE UNIVERSITY OF ALBERTA

RELEASE FORM

NAME OF AUTHOR Elizabeth T. Nadon

TITLE OF THESIS Early Indices of Grade One Academic Performance

DEGREE FOR WHICH THESIS WAS PRESENTED Master of Education

YEAR THIS DEGREE GRANTED Fall, 1983

Permission is hereby granted to THE UNIVERSITY OF ALBERTA LIBRARY to reproduce single copies of this thesis and to lend or sell such copies for private, scholarly or scientific research purposes only.

The author reserves other publication rights, and neither the thesis nor extensive extracts from it may be printed or otherwise reproduced without the author's written permission.

(Signed) Elizabeth Nadon

PERMANENT ADDRESS:

18 Bluebird Crescent

Sherwood Park, Alberta

T8A 0E7

DATED June 22/83

THE UNIVERSITY OF ALBERTA

FACULTY OF GRADUATE STUDIES AND RESEARCH

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research, for acceptance, a thesis entitled Early Indices of Grade One Academic Performance submitted by Elizabeth T. Nadon in partial fulfilment of the requirements for the degree of Master of Education in Educational Psychology.

.....
Supervisor

.....

.....

Date..June 3/83.....

## ABSTRACT

The present study was undertaken to investigate the relationship between childrens' kindergarten performance and their grade one general academic standing. The early assessment instruments included three tests of visual-motor functioning (The Developmental Test of Visual-Motor Integration, The Draw a Person Test, and the Group Perceptual Screening Test), and a teacher rating scale (Myklebust's Pupil Rating Scale). In addition the variables of sex, age and daycare (CARE) were considered. As a sub-objective of the present research was to examine the efficacy of using the Metropolitan Readiness Tests (administered at the beginning of grade one) in making placement decisions about children, the composite score from these tests was included in the study and employed in a dual function as both a predictor and a criterion variable. The constructs underlying these instruments and the rationale for their inclusion were discussed in terms of the previous research.

Subjects were 204 children registered in English language ECS (i.e. kindergarten) programs in the County of Strathcona. This represented the total population of English ECS children enrolled in four schools in Sherwood Park. After the one year period it was possible to obtain follow-up data on 84.3% of the original group (172/204).

The statistical procedures employed in this study included

a descriptive analysis of the data in terms of means, standard deviations, and intercorrelations (for the total group and separately by sex), and separate stepwise multiple regression analyses using the composite score of the MRT and general grade one academic standing as criterion variables. Finally, classificational analyses using each of the predictor variables in isolation were undertaken to determine where prediction errors would occur given a specific set of criteria.

The results of the descriptive analysis revealed, first of all, that girls achieved higher scores than boys on most of the variables and, secondly, that the correlations between predictor and criterion variables were somewhat stronger for girls than for boys. The results of the multiple regression analysis (MRT as criterion) indicated that only two predictor variables, the VMI and the PRS contributed significantly to the equation when the group was considered as a whole. When boys and girls were considered separately, the daycare variable (CARE) contributed significantly to the multiple correlation coefficient. When grade one academic standing was the criterion, the order in which the variables entered the equation (MRT, PRS, VMI) remained invariant for the group as a whole and the subgroups (boys and girls). In the case of the boys, however, the VMI did not contribute significantly to the multiple correlation coefficient. The classificational analyses revealed that, of the five pre-

dictors, the PRS and the MRT were about equal in predicting grade one academic standing. Using either of these instruments resulted in the identification of approximately 55% of the 'at risk' pupils with less than 15% of the 'not at risk' pupils being screened in by the procedure.

As a result of the present study it was concluded that the use of any of the predictor variables under consideration either alone or in combination for the purpose of making placement decisions about individual children was unwarranted. The efficacy of looking for accurate predictors of academic achievement among a preschool group of children was discussed and suggestions were made as to some directions which future research might take.

Acknowledgements

I would like to express my sincere thanks to all those people who have assisted me in a number of ways over the past several years and without whom this thesis would not have been completed. I am especially indebted to Dr. E. Fox whose encouragement and support gave me the confidence to enter the graduate program in the first place.

I wish to thank the County of Strathcona School Board for allowing me to conduct my study in county schools. I owe a special thanks to the administrative staff, counsellors, kindergarten and grade one teachers of Brentwood, Campbelltown, Glen Allan, Mills Haven, and Wes Hosford Elementary Schools who gave so generously of their time and energy in order to carry out the testing and complete the numerous forms within the allotted time frame. This study could not have been completed without their cooperation.

I am extremely grateful to Dr. H. Janzen, my Thesis Supervisor, for his direction and many helpful suggestions and for his uncanny knack of knowing just when to offer encouragement. I would also like to thank Drs. F. Boersma and J. Blakey for so generously agreeing to serve on my Committee at a very busy time of year.

My final thanks goes to my husband, Dan, and my children, Rochelle and Jeannette, who managed to endure my obsessive work habits for almost two years with grace and good humor. Their constant support, understanding and tolerance contributed immeasurably to the completion of this project.

TABLE OF CONTENTS


| CHAPTER | | PAGE |
|---|---|---|
| 1. | STATEMENT OF THE PROBLEM | 1 |
| 2. | REVIEW OF THE LITERATURE | 8 |
| | Teacher Evaluations | 9 |
| | Readiness Tests | 23 |
| | Psychometric and Combined Approaches | 31 |
| 3. | DESIGN AND PROCEDURES | 52 |
| | Subjects | 52 |
| | Procedure | 53 |
| | Instrumentation | 57 |
| | Draw a Person Test | 59 |
| | Developmental Test of Visual-Motor Integration | 61 |
| | Group Perceptual Screening Test | 62 |
| | Pupil Rating Scale | 63 |
| | Metropolitan Readiness Tests | 65 |
| | Data Analyses | 67 |
| 4. | RESULTS | 68 |
| | Descriptive Statistics | 69 |
| | Multiple Regression Analyses | 72 |
| | Classificational Analyses | 78 |
| 5. | DISCUSSION AND CONCLUSIONS | 85 |
| | Discussion and Implications | 86 |
| | Recommendations and Suggestions for Future Research | 95 |

CHAPTER PAGE

BIBLIOGRAPHY..... 99

APPENDIX A..... 110

APPENDIX B..... 114

Appendix C..... 116

LIST OF TABLES


| Table | | Page |
|---|---|---|
| 1 | Prediction of Grade One Reading Performance Using the SRS | 21 |
| 2 | Prediction of Third Grade Reading Performance Using the SRS | 21 |
| 3 | Prediction of Overall Grade One Performance as Determined by Teacher Rating and Scores on the CAT Using the MRT | 26 |
| 4 | Prediction of Overall Grade One Performance Using the L-C Reading Readiness Test | 26 |
| 5 | Prediction of Overall Grade One Performance Using the First Grade Screening Test | 29 |
| 6 | Prediction of Grade Two Reading Performance Using the Jansky and deHirsch Screening Index | 37 |
| 7 | Prediction of Grade One Reading Performance on the Gates-MacGinite Reading Test Using the deHirsch Predictive Index | 39 |
| 8 | Prediction of Grade Two Achievement as Exemplified by Teacher Recommented Grade Placements Using a Modified Predictive Index | 39 |
| 9 | Prediction of Grade Two Reading Level Based on the Kindergarten Discriminant Function Composite Scores | 42 |
| 10 | Prediction of Second Grade Reading Achievement Based on Kindergarten Group Designation | 46 |
| 11 | Prediction of Second Grade Reading Achievement Based on Kindergarten Group Designation | 46 |
| 12 | Means and Standard Deviations of all Variables | 70 |
| 13 | Differences in Means (Boys versus Girls) | 71 |
| 14 | Correlation Matrix for All Variables (Total Group) | 73 |

| Table | | Page |
|---|---|---|
| 15 | Correlation Matrix for All Variables (Girls) | 74 |
| 16 | Correlation Matrix for All Variables (Boys) | 75 |
| 17 | Multiple Regression with Metropolitan Readiness Test (MRT) Total as Criterion | 77 |
| 18 | Multiple Regression with Grade One General Academic Standing as Criterion | 79 |
| 19 | Prediction of Grade One Academic Standing Using the Pupil Rating Scale | 82 |
| 20 | Prediction of Grade One Academic Standing Using the Draw a Person Test | 82 |
| 21 | Prediction of Grade One Academic Standing Using the Group Perceptual Screening Test | 83 |
| 22 | Prediction of Grade One Academic Standing Using the Test of Visual-Motor Integration | 83 |
| 23 | Prediction of Grade One Academic Standing Using the Metropolitan Readiness Test | 84 |

# CHAPTER 1

## Statement of the Problem

Educators have always been concerned with the problem of prediction. At the end of each school year teachers must make a decision as to whether or not a child is ready to proceed to the next grade. Historically this decision was simply based upon the child's mastering or failing to master some predetermined amount of material as measured by teacher-made tests. This approach gradually gave way to the pressures of equality of educational opportunity, the establishment of regional norms and standardized testing procedures which made the decision as to whether or not a child was ready for the next grade more remote from the actual classroom curriculum content and introduced an additional element of uncertainty into the decision making process. With the advent of universal kindergarten programs in Alberta, the problem of prediction has been shifted down a year. Not only is there no real academic content to measure but developmentally the child is in a state of flux (Flynn & Flynn, 1978) making standardized testing procedures far from reliable.

In addition to the actual problems of assessment there has been a growing awareness of the negative psychological consequences which might ensue should children be placed in academic settings beyond their level of coping (Raccioppi, 1982).

There has been increasing pressure to identify children who are likely to experience academic problems before failure occurs in the hope of avoiding the negative consequences incurred by constant exposure to discouraging and frustrating situations. If these children could be identified, early intervention could possibly alleviate years of frustration and prevent ensuing social and emotional problems (Jansky & de Hirsch, 1972; Lindsay & Wedell, 1982; Raccioppi, 1982).

Although educators for years have been concerned with the accurate prediction of school achievement, the idea that failure could be prevented has only recently come into vogue. Indeed a review of the literature concerning learning problems over the last two decades reveals a shift in focus from a medical-treatment model (i.e. diagnose the problem and then institute a remedial program to "cure" it) to a prevention model (i.e. identify the precursors of learning difficulties and intervene before a problem develops). Friedman, Fuerth, and Forsythe (1980) attribute this shift in attitude largely to the appearance in 1966 of de Hirsch, Jansky and Langford's work entitled Predicting Reading Failure. This work rests on the assumption that if we can predict reading failure we can prevent it and stimulated interest in the test battery approach to prediction. Since that time a major growth area in educational research has been the early identification of children likely to experience learning difficulties. Maitland, Nadeau, and Nadeau (1974) cite

the explosion in the number of readiness tests on the market as evidence of this increased emphasis on early detection. They note that only eight readiness tests are listed in the sixth edition of Buros Mental Measurements Yearbook (1965) whereas twenty-nine are listed in the seventh edition (1972). Lindsay and Wedell (1982) note that this trend reached England in the 1970s. Surveys there showed that the percentage of local education authorities using early screening procedures increased from 25% in 1972 to 47% in 1976 and that a further 27% were planning to implement them. Unfortunately, as these authors note, the movement to adopt some type of formalized early screening procedure has not been matched by a show of equal enthusiasm in evaluating the effectiveness of these procedures.

The emphasis of the present study is to evaluate the efficiency of some of the procedures used by this researcher in arriving at a decision about whether or not a child is ready to begin formal schooling (i.e. enter first grade). The instruments selected and their rationale for inclusion in the study will be discussed in chapter 3.

Before looking at the literature related to this topic it is important to clarify what it is that we are trying to predict and subsequently prevent. The research reflects three seemingly divergent focii: 1) prediction of reading failure, 2) prediction of school failure per se, and 3) prediction of dyslexia. In line with these "different" approaches a variety

of measurement instruments have been produced which according to Evans and Ferguson (outlined in Lindsay and Wedell, 1982) can be classified as: 1) reading readiness, 2) general academic readiness, and 3) identification of learning disabilities or potential. Lindsay and Wedell point out, however, that the instruments involved are all remarkably similar in content. They are generally divided into five or six subtests which purport to measure some aspect of the child's development such as language, visual and auditory perception, motor skills, perceptuomotor functioning and letter recognition. The terms reading failure and school failure are closely related and often used interchangeably in the literature. Dyslexia and learning disabilities, however, have a somewhat different connotation and generally implicate basic psychological and/or neurological processes. For a review of the complexities involved in defining these terms see Rutter (1978), Eisenberg (1978) and Mattis (1978). In any case, whether the researchers set out to predict school failure, reading failure or dyslexia, because of the similarity of the devices used they are likely identifying the same subgroup of children at least initially.

Silver (1978) considers that prediction and subsequently prevention can be regarded as either "primary" or "secondary". Primary prevention involves identifying the root causes of reading failure whether they be biological, psychological, social or educational and implementing appropriate intervention

programs. In terms of this model prenatal parental education programs, nutrition, well-baby clinics, and early childhood programs (such as the Headstart project) would be considered primary preventative programs. In contrast, secondary prevention refers to strategies designed to meet the needs of a select group of children for the specific purpose of circumventing predicted academic difficulties. As a first step, secondary prevention requires, according to Silver (1978), that the school system (or other agency) "detect independently of cause those children likely to fail in reading" (p. 353). Once identified, these children would be further evaluated and specific strategies would be designed to meet their individual needs. This three step process is referred to as scanning, diagnosis and intervention. The present study as well as the studies reviewed in the following chapter are concerned with only the first step in this process. In other words, the focus is on isolating at the kindergarten level those children who are likely to experience academic difficulties unless they receive some type of specialized assistance at an early stage in their school career. At the point of secondary prevention the issue of causality is no longer of primary importance. It is recognized that the factors underlying each individual child's difficulties may be many and varied. Silver stated the purpose of preschool screening clearly and succinctly:

> Scanning is designed only to detect a heterogeneous group whose members have but one factor in common:

> their function in those parameters assessed by the scanning instrument is immature when compared to their peers. (354)

Belmont and Belmont (1980) in their studies of reading achievement have identified four subgroups of children who exhibit learning problems: 1) those who will be permanently disabled learners, 2) those who perform poorly early in their school career and later catch up, 3) those who will be adequate early and poor later and 4) those who will fluctuate over the years between failure and acceptable performance.[1.] These researchers consider the first group of children to be learning disabled, the second to be developmentally delayed and the fourth to be simply low normal children who are exhibiting expected fluctuations in level over time. They suggest that children belonging to the third group may be learning disabled for more complex cognitive reasons but other explanations such as inadequate teaching, poor motivation, failing health or complex social or emotional problems might also pertain to members of this group. Applying this frame of reference to the present study, it is predicted that the group of "at risk" children identified by the screening procedures will encompass those children who are, in fact, learning disabled, those who are developmentally delayed and some of those who belong to the low average goup. As both predictor and criterion measures are

1. Satz, Taylor, Friel and Fletcher (1979) report similar findings as a result of their six-year longitudinal study.

to be administered at an early stage in the academic experience of these children it is recognized that group three, those children who fail after an initial success will be missed entirely.

## CHAPTER 2

## Review of the Literature

A number of researchers (Lindsay & Wedell, 1982; Maitland, Nadeau & Nadeau, 1974) make reference to the recent proliferation of instruments and assessment procedures used for the prediction of academic performance. They lament not only the paucity of technical data related to the validity and reliability of many of the instruments now available and in use but the fact that follow-up evaluations of the procedures used are few and far between.

The review of the literature which follows will focus on research undertaken during the last two decades which pertains to the prediction of early adademic performance. The chapter will be divided into three subsections each of which will represent a specific procedural approach: 1) predictions based on teacher evaluations, 2) predictions based on readiness tests, and 3) predictions based on psychometric evaluations and combined techniques. Since a specific concern of the present study is to ascertain how much weight might reliably be given to a variety of commonly used tests and assessment procedures when we are asked to make programming decisions about individual children, a major focus of this chapter will be to assess, whenever possible, the accuracy of prediction of the specific procedure under review as determined by the number of children

correctly identified as belonging either to the failure or to the success group.

Teacher Evaluations

It makes intuitive sense to expect teachers, on the basis of their familiarity with their students in their day-to-day performance, to be able to predict how they will perform in subsequent years. The teacher is familiar with the expectations of the school and is experienced in viewing children in the academic setting. As Keogh, Tchir and Windeguth-Behn (1974) observe:

> The kindergarten and primary grade teacher's day-to-day experience with a variety of behaviors gives her an unequaled perspective for appraising inappropriate or deviant behaviors. She is probably the first professional to observe and compare a child with his peers. (373)

Teacher's predictions then are based on their unique position and training. They have the opportunity to ascertain the child's present performance, they are in a position to judge that performance in relation to that of many other children, and they know the academic expectations and the tolerance level of the school system in which they work.

In a concurrent validity study involving 276 first grade children Kermonian (1962) found that teacher ratings after only two weeks of exposure to their students on the dimensions of reading readiness, number readiness and total readiness correlated .73, .73 and .77 respectively with the same dimensions

as measured by the Metropolitan Readiness Test. Although teachers tended to rate higher than the MRT, 85.3% of all teacher estimates fell within one rank of the formal test results. No attempt was made to ascertain whether the teacher ratings or the readiness test results were more indicative of future achievement.

One aspect of a study by Keogh and Smith (1970) involving an initial sample of 59 kindergarten children examined the efficacy of using a simple teacher rating as a predictor variable of achievement in grades two through five. They found significant correlations between teacher ratings and later achievement measures at all levels and for both boys and girls. This study revealed an interesting sex difference in that the teacher rating was more highly correlated with girls' math achievement (r=.73) as opposed to girls' reading achievement (r=.59) whereas the opposite pattern was noted for boys. The average correlation between the teacher rating and the boys' math achievement was .42 whereas the average correlation with the boys' reading achievement was .56. Overall there was a stronger relationship between teacher rating and the girls' performance.

Telegdy (1975) attempted to assess whether teacher ratings or readiness test results were more predictive of academic achievement. He included teacher predictions along with a number of readiness tests in a study (N=56) designed to assess

the predictive validity of each of the measures in terms of its correlation with academic performance at the end of the first grade as measured by the Wide Range Achievement Test and Gray's Oral Reading Test. The correlations between teacher predictions and the criterion variables ranged from .35 to .65, considerably lower than those reported between the Metropolitan Readiness Test and the criterion variables which ranged from .58 to .73. This study supported the notion of using standardized testing procedures in conjunction with teacher observations as a basis for predicting academic performance.

A somewhat more sophisticated method of achieving teacher input into early detection of learning difficulties is the formalized rating scale. Rather than assigning a child a single rating on the basis of overall academic readiness, for example, the rating scale generally comprised a number of items grouped into subtests each purporting to measure some specific aspect of the child's present functioning. Theoretically, because the teacher must focus on a number of discreet dimensions, the rating scale should allow for a more objective evaluation of each child. Stevenson, Parker, Wilkinson, Hegion, and Fish (1976), however, have questioned the need for lengthy rating scales at the kindergarten level. In a study involving 217 kindergarten children followed through to the end of grade three they found that no combination of three or more ratings predicted achievement significantly better than the best one or two ratings. On the

basis of a correlational analysis they claimed that four variables taken together (Effective Learning, Vocabulary, Following Instructions, and Retaining Information) provided an efficient scale which, although somewhat redundant at times, could be used at every grade level. The correlations between the sum of these four ratings completed in the spring of the kindergarten year and achievement scores at the end of grade one, two, and three were .52, .61, and .48 respectively. Although these correlations were statistically significant, they accounted for only a small percentage of the total variance (roughly 29%). No attempt was made to assess the efficiency of using this procedure in terms of the number of individual children identified who later experienced reading difficulty.

An example of a formalized rating scale which has been used with kindergarten children is the Pupil Rating Scale (PRS) (Myklebust, 1971). This scale consists of twenty-four items which are grouped into five subscales: auditory comprehension, spoken language, personal-social behavior, and motor coordination. Each item is rated on a scale of one to five with lower scores reflecting deficits. Developed for somewhat older elementary school children (it was normed on third and fourth graders) for the purpose of identifying children with learning disabilities, its appropriateness for use with kindergarten pupils has only recently been investigated. Federici, Sims, and Bashian (1976) conducted a study in which high and low risk kindergarten children,

as determined by their performance on the Meeting Street School Screening Test (MSSST), were subsequently rated by teachers using the PRS. The resulting significant difference in means between the two groups attested to the validity of using the PRS at the kindergarten level as a means of identifying "at risk" children. The correlation between the PRS and the MSSST for the high risk sample, however, was only .39, indicating that the complexion of the group might vary considerably depending upon the instrument used to define it. No attempt was made to assess the predictive validity of either instrument in terms of later academic achievement.

In another concurrent validity study Colligan (1979) examined the relationships between the PRS and several criterion variables: the Lippincott Readiness Test, the Metropolitan Readiness Test and the Wide Range Achievement Test. Substantial correlations ranging from .63 to .77 were reported. In addition a correlation of .66 was noted between a low PRS score and a child's participation in a remedial program. Two years later, at the end of grade two, in an attempt to assess the predictive validity of the PRS, Colligan (1979) collected achievement data on this same group of children (N=55). In addition to three subtests of the Stanford Achievement Test (Paragraphs, Word-Study, and Math) a disability rating based on the amount of supplemental instruction the child had received over the three years was included as a criterion variable. The results of the correlational analysis

indicated highly significant relationships ranging from .53 to .72 between the PRS scores and the criterion measures. The Auditory Comprehension subscale was most highly related with all the criterion variables except Math (r=.67) which showed a slightly stronger association with the Spoken Language subscale (r=.71). Colligan also examined the means and standard deviations of the total kindergarten study group and each of the subgroups (based on the degree of reading disability) and compared these figures with those from Myklebust's original study involving third and fourth grade pupils. A very close correspondence in mean scores was noted between Colligan's kindergarten study group which had the greatest difficulty in learning to read (i.e. the group which had had supplemental assistance for all three years) and Myklebust's learning disabled group. On the basis of this study, Colligan concluded that the PRS was suitable for use with a kindergarten population and possessed sufficient predictive validity to warrant its inclusion in any battery of tests designed to identify children who might experience future learning difficulties.

Another study concerning the utility of the PRS was conducted by Margolis, Sheridan and Lemanowicz (1981) using grade one and two pupils. In contrast to the studies reviewed thus far, this group of researchers carried out a classificational analysis of the data as well as the more traditional correlational analysis. Although significant correlation coefficients ranging

from .50 to .68 between the PRS and the Cooperative Primary Tests administered concurrently were reported, the classificational analysis was less impressive. Of the grade one children designated as "at risk" on the basis of Myklebust's critical cutoff score for suspected learning disabilities (total PRS score $\leq$ 66) only 17%, 6% and 17% scored in the lower quartile for Word Analysis, Arithmetic and Reading respectively. Although there was a stronger trend for correctly identifying grade two children with reading and arithmetic problems (27%, 39% and 33% respectively), the majority of the children identified by the scale as being at risk for possible learning disabilities did not experience academic difficulties.

Another formalized rating scale which has been used with kindergarten populations and has undergone rather intensive study is the Student Rating Scale (SRS) developed by Adelman and Feshbach (1971). This scale consists of forty-one items and samples aspects of the child's cognitive, affective and social functioning in the classroom. Like the PRS, each item is rated on a scale of one to five with lower scores reflecting deficits.

In 1972 Feshbach, Adelman and Fuller (1974, 1977) began a study involving two large cohorts of kindergarten children (N=888 and 844 respectively) which set out to compare the relative effectiveness of using the SRS as apposed to psychometric procedures in predicting future academic achievement.

The results of the correlational analysis at the end of the first grade indicated that using z-scores based on within-class means and standard deviations the SRS total score and the SRS Factor III (visual and auditory perceptual discrimination) yielded the highest correlations of all the predictor variables.[1] In the Year I sample (N=433) these two variables correlated .60 with the Gates MacGinite raw score and .50 and .51 respectively with the Cooperative Primary Test raw scores. The SRS total score emerged as the best single predictor in the multiple regression analysis as well.

The follow-up conducted for both cohorts of children (N=403 and 464 respectively) at the end of the third grade revealed the presence of sex differences. Girls consistently achieved higher scores on each of the achievement tests administered. As a result correlational analyses were carried out separately by sex as well as for the total group. The SRS still emerged as the best predictor overall but some interesting sex differences were revealed: IQ was more highly correlated with achievement for boys than for girls across all grade levels; the SRS total and all factor scores except language were more highly correlated with first grade achievement scores for boys than

1. Feshbach et al. (1977) contend that although the use of z-scores based on within-class means and standard deviations may result in a loss in discrimination entailed in assuming equivalence between classes this is more than made up by the reduction of error variance brought about by the tendency of some teachers to avoid low ratings.

for girls whereas by the third grade the correlations were higher for girls. The authors suggest that motivational variables may be more important for girls in the early primary grades and generally outweigh the influence of IQ. The kindergarten behavioral factors, on the other hand, which generally reflect poorer performance for males than for females likely decline in importance for males with increasing maturity. Of the SRS factor scores the best predictor for boys across all grade levels was Factor II (language) with correlations of .41, .40, and .37 with first, second, and third grade reading achievement respectively. For girls the best predictor was Factor V (perceptual-motor) with correlations of .37, .45, and .32 with first, second, and third grade reading achievement.

It is one thing to show a relationship between an instrument administered in kindergarten and later academic performance and quite another to predict on the basis of that first test which individual children are likely to fail in reading at the later date. There are a number of factors which make such a prediction hazardous. Keogh (1971) stresses that what children do is more predictive than what they do not do. High performance on a task is a demonstration of proficiency whereas poor performance does not necessarily indicate lack of proficiency. It may be attributable to a number of other factors such as lack of motivation, poor eyesight, inability to follow the directions, anger, or fatigue, to mention a few. Feshbach et al. (1977) also point to

such intervening variables as the child's home and school environment and the variable rate of development characteristic of 5 to 7 year old children as having an effect on his achievement. Using the mean reading achievement scores of the first grade classrooms as one index of first grade experience and the initial kindergarten Student Rating Scores as the other variable they demonstrated that second grade achievement scores varied both as a function of the initial SRS score ($p<.001$) and as a function of the first grade experience ($p<.001$). In other words, children who received a low SRS score in kindergarten, not surprisingly, achieved better reading scores in grade two if they had been in the higher achieving grade one class. Those children who had received average SRS scores in kindergarten were even more markedly affected by the grade one experience than either the low or the high SRS group. The data from this study also revealed that the percentage of children predicted to succeed but who were performing below grade level varied as a function of the school they attended.

Although the early identification of children at risk for reading failure is fraught with inherent difficulties, the need to identify these children so that early intervention procedures can be instituted remains. In an attempt to assess the predictive efficiency of the SRS as a screening instrument, Feshbach et al. (1974, 1977) carried out classificational analyses of their data. In their one-year follow-up, using a z-score of -1.25 on the SRS

as the prediction cutoff score and a grade score of less than 1.7 as the criterion of reading failure, they reported a false positive rate (i.e. children identified by the test as "at risk" who did not have any future difficulties) of 17% and a false negative rate (children who later had difficulties but were not identified by the test) of 24%.

In a critique of the Feshbach et al. (1974) study, Satz and Fletcher (1979) point to methodological weaknesses in their organization of the detection tables. Based on the earlier work of Meehl and Rosen (1955) these researchers and others (Mercer, Algozzine, & Trifiletti, 1979) argue that the rows of a detection table must represent the predictor variable and the columns the criterion outcomes. Organized this way Mercer et al. make provision for both vertical and horizontal analyses whereas Satz and Fletcher stress that any discussion of the percentage of "hits" and "misses" must relate to the vertical analysis only. The horizontal analysis yields information as to the probability of accurate designation if a child is labled "at risk" or "not at risk" by the predictor. Failure to adhere to a consistent definition of these terms has led to much confusion in the literature and inappropriate use of screening instruments. In order to maintain clarity and facilitate comparative analyses with other studies, the data from the Feshbach et al. research will be reorganized and presented according to the format outlined above as will the data from all other

studies reviewed here.

Looking, first of all, at the preliminary follow-up study, the data reveal that the SRS identified 54 of the 179 (30.2%) children who later experienced reading difficulty and 395 of the 406 (97.3%) who later were reading at an adequate level (see Table 1). It can readily be seen that the number of children incorrectly labled as "at risk" (i.e. the false positives) was very small (11/406 or 2.7%). The false negative rate (those children who later experienced difficulty but were not identified by the test), on the other hand, was extremely high (125/179 or 69.8%). An analysis of the data from the end of grade three follow-up study reveals a similar pattern although the percentage of correctly identified "at risk" children is considerably lower (see Table 2). Although the overall hit rate in both of these analyses seems quite impressive (77% and 70% respectively), it is evident from the high false negative rates that the SRS is not an efficient predictor of future learning difficulties--the task for which it was designed.

From the research reviewed in the foregoing section it would seem that the results could best be described as equivocal. It is evident that significant correlations in the neighborhood of .40 to .70 between teacher predictions and future academic performance are quite common whether these predictions are based of a simple rating of general academic readiness or a complex rating scale. What we should do with this information

Table 1

Prediction of Grade One Reading Performance Using the SRS*

| SRS | Gates MacGinite Reading Test | | |
|---|---|---|---|
| | Poor (≤1.7) | Good (>1.7) | Total |
| Poor (z ≤ -1.25) | 54 (30.2%) | 11 (2.7%) | 65 |
| Good (z > -1.25) | 125 (69.8%) | 395 (97.3%) | 520 |
| Total | 179 | 406 | 585 |

Total hits = 449/585 (76.8%) Total misses = 136/585 (23.2%)

*Data from Feshbach, Adelman, and Fuller (1974).

Table 2

Prediction of Third Grade Reading Performance Using the SRS*

| SRS | Cooperative Reading Test | | |
|---|---|---|---|
| | Below grade level (≤ 3.7) | At grade level (> 3.8) | Total |
| Poor (z ≤ -1.25) | 26 (17.7%) | 13 (4.3%) | 39 |
| Good (z > -1.25) | 121 (82.3%) | 291 (95.7%) | 412 |
| Total | 147 | 304 | 451 |

Total hits = 317/451 (70.3%) Total misses = 134/451 (29.7%)

*Data from Feshbach, Adelman, and Fuller (1977).

is less clear. Some research suggests that teachers may rate girls more accurately than boys. Another study reflects the opposite trend with the correlation between teacher prediction and performance stronger for boys than for girls, at least at the end of the first grade. One fact, however, is clear. A significant correlation between predictor variable and criterion outcome does not imply that a screening test or procedure is suitable for making decisions about individual children. Classificational analyses reviewed here have generally shown that the instruments failed to identify many children who later experienced difficulties. Although manipulating the cutoff scores of the predictor and criterion variables might alleviate this problem to some extent, the real source of the difficulty would seem to be that teachers may be overly optimistic in their predictions of future success and that they may hesitate to forecast future disability. If, however, a child is rated as likely to experience difficulty the chances seem to be about three out of four that he will indeed have problems.

In terms of the present study, it is predicted that there will be a significant correlation between teacher rating and future academic achievement. Furthermore it is predicted that the teacher evaluations will be a significant factor in the multiple regression analysis.

Readiness Tests

The readiness test approach to prediction rests upon the widely accepted notion that present performance is predictive of future achievement. As children at the preschool level cannot yet read, write or do mathematical operations, however, assessing their present level of functioning in terms of these skills presents a major problem. The general solution to this dilemma has been to design test items which measure skills thought to be precursors of the target task. The predictive validity of the test, then, rests upon two factors: 1) the degree to which the selected antecedent task is in fact related to the target task and 2) the degree to which present performance actually reflects future achievement. With regard to the first factor it would seem reasonable to predict that a child who has already mastered a skill such as phoneme-grapheme association or recognizes and can isolate initial consonant sounds or has the patience to check out a number of abstract visual symbols in a matching task is more likely to read sooner than a child who has not yet mastered any of these skills. The importance of the second factor, although certainly subject to the influence of many intervening variables, has generally been supported in the literature (Friedman, Fuerth, & Forsythe, 1980).

In a study involving 2,069 pupils, Bremer (1959) reported a correlation of .40 (significant at the .01 level) between the Metropolitan Readiness Test (MRT) administered at the beginning

of grade one and achievement measured one year later at the beginning of grade two. Similar results were reported by Stevenson, Parker, Wilkinson, Hegion and Fish (1976). In their study involving an initial sample of 255 children they found the correlation between the total score on the MRT administered at the beginning of grade one and the reading subtest of the Wide Range Achievement Test (WRAT) administered at the end of grade one to be .44 for boys and .48 for girls.

In addition to the correlational analysis Bremer also carried out a classificational analysis of his data. He discovered that only a third of those children who scored low in readiness actually achieved at the lowest level in reading whereas an equal number achieved at an average or above average level. He concluded from his study that in spite of the statistically significant relationship between the MRT and later achievement, using readiness test scores as an aide to prediction would result in only a slight increase in accuracy. He noted that some children seemed to overcome deficits in readiness more easily than others and that, rather than for placement decisions, readiness tests should only be used in a diagnostic sense to set up extremely flexible and temporary within class remedial groupings.

In contrast to the two foregoing studies, Lessler and Bridges (1973) reported a correlation of .76 between the MRT total score and the end of grade one performance as measured by the California Achievement Test (CAT). Their study, involving

a sample of 293 first grade children, compared the effectiveness of using a number of variables both alone and in combination as predictors of future achievement. In addition to the MRT the variables included the Lee-Clark Reading Readiness Test (L-C), the California Test of Mental Maturity (CTMM), the Peabody Picture Vocabulary Test (PPVT), and the Bender Gestalt Test (BG). The classificational analysis of their data revealed that the MRT alone correctly identified 129 (87.2%) of the 148 children who later exhibited learning problems and 124 (85.5%) of the 145 children who later achieved adequately for a total hit rate of 86.3% (see Table 3). Using the L-C was somewhat less effective with a true positive rate of 77% and a true negative rate of 82.8% (see Table 4). On the basis of this research both the MRT and the L-C would seem to be valid screening devices on which to base decisions with regard to group placement and remedial intervention.

In addition to the relationship between teacher ratings and future achievement reported in the preceding section, Telegdy (1975) also investigated the relationship of three readiness tests with first grade achievement. In addition to the MRT the Screening Test of Academic Readiness (STAR) and the First Grade Screening Test (FGST) were used as independent variables. Although significant correlations were found between all predictor and criterion variables, the MRT total score was more highly correlated with the criterion variables than were the

Table 3

Prediction of Overall Grade One Performance as Determined by Teacher Rating and Scores on the CAT Using the MRT*

| MRT | Grade One Performance | | |
|---|---|---|---|
| | Poor | Adequate | Total |
| Poor | 129 (87.2%) | 21 (14.5%) | 150 |
| Adequate | 19 (12.8%) | 124 (85.5%) | 143 |
| Total | 148 | 145 | 293 |

Total hits = 253/293 (86.3%) Total misses = 40/293 (13.7%

*Data from Lessler and Bridges (1973).

Table 4

Prediction of Overall Grade One Performance Using the L-C Reading Readiness Test.*

| L-C | Grade One Performance | | |
|---|---|---|---|
| | Poor | Adequate | Total |
| Poor | 114 (77%) | 25 (17.2%) | 139 |
| Adequate | 34 (23%) | 120 (82.8%) | 154 |
| Total | 148 | 145 | 293 |

Total hits = 234/293 (79.9%) Total misses = 59/293 (20.1%)

*Data from Lessler and Bridges (1973).

total scores of the other two readiness tests. Similar to the Lessler and Bridges study correlation coefficients ranging from .58 to .73 were reported. The highest correlation between predictor and criterion variables, however, was reported for the Letters subtest of the STAR with correlation coefficients ranging from .62 to .76. Indeed, using the total scores of all the predictor variables in multiple correlation did not add to the predictive validity of using this one subtest in isolation (R's ranged from .64 to .74).

In a retroactive study involving 28 third grade subjects and designed as a follow-up to the work of Telegdy and others (Hutton, 1970;Klein, 1977) Nichta, Federicic and Schuerger (1982) investigated the relationship between the STAR administered prior to kindergarten admission and grade three achievement as measured by the Peabody Individual Achievement Test (PIAT). Significant correlations ranging from .40 to .71 were reported between the STAR total score and the age equivalent scores of the PIAT.

In a study involving a relatively large sample of kindergarten children (N=212) Hase (1977) found the FGST to be an efficient predictor of first grade performance. In addition to the total score this researcher grouped the test items into four subtests (Self-Concept, Visual-Motor, Reasoning, Vocabulary) and used these scores as well in identifying the "at risk" group. The procedure used was to consider a child "at risk" for poor school success if his total score was 15 or less (the cutoff

score suggested by the test authors) or if he scored in the moderate or high risk range (scores established in an earlier pilot study) on any of the four subscales. In a comparison of the test classification of "at risk" verses "not at risk" and the teacher ratings of "poor" verses "adequate" in overall school performance, agreement was found in 80% of the cases, very similar to the overall hit rate reported by Lessler and Bridges using the MRT and the L-C. A closer examination of the data provided by Hase, however, yields a somewhat less favorable picture (see Table 5). Although this test correctly identified 157 out of 189 children who later achieved adequately (true negative rate = 82.8%), it identified only 17 out of 23 children who later experienced difficulty (true positive rate = 74%). In other words it missed 26% (roughly 1 in 4) of those children who later had academic difficulties and screened in 17% of the group of children who later had no difficulties. A horizontal analysis of the data reveals that of those children predicted to do poorly (N=49) only 17 (34%) actually had subsequent difficulty. In other words, even if an individual child was designated as "at risk" by this test, the likelihood of having future problems would only be about 1 in 3. If, on the other hand, a child was designated as "not at risk" by this test the chances of not having any future problems would be better than 9 out of 10. Although the FGST would seem to be an efficient predictor of success based on these figures it can readily be seen that using no test at all would yield similar

Table 5

Prediction of Overall Grade One Performance using the First Grade Screening Test.*

| | Teacher Rating of Grade One Standing | | |
|---|---|---|---|
| FGST | Poor | Adequate | Total |
| At Risk | 17 (74%) | 32 (17%) | 49 |
| Not at Risk | 6 (26%) | 157 (83%) | 163 |
| Total | 23 | 189 | 212 |

Total hits = 174/212 (82%) Total misses = 38/212 (18%)

*Data from Hase (1977).

results. If we simply predicted that all of the children would achieve adequately, we would be correct in 189 out of 212 cases (89% total hit rate) and we would be incorrect in only 23 (11%) cases. It is, however, precisely this group which we are attempting to identify. Using no test at all is simply not a viable solution as this course of action would result in a false negative rate of 100%.

As in the preceding section, conclusions as to the utility of using readiness tests as predictive of future academic achievement must remain tentative. All studies reviewed here reported significant positive correlations between readiness test performance and subsequent academic achievement but these correlations varied greatly indicating wide fluctuations in predictive utility. The classificational analyses, too, did not yield uniformly positive results although the MRT, in terms of the proportion of children with learning problems who were identified (true negatives), the proportion of "at risk" children who in fact later experienced difficulty, and the total overall hit rate, generally seemed to be an effective predictor of future performance.

In terms of this study, it is predicted that there will be a significant correlation between performance on the MRT administered in the fall and general academic performance as determined by teacher ratings later in the year. Furthermore, it is predicted that the readiness test results will contribute significantly to the correct classification of "at risk" verses "not at risk" students.

Psychometric and Combined Approaches

In contrast to predictions based on teacher ratings and readiness tests which basically rank children on presumed pre-academic skills within their peer group, whether immediate or more broadly based, psychometric approaches to prediction presume that an understanding of the child in terms of overall physical and cognitive development will yield data predictive of future functioning in academic tasks. There are two opposing schools of thought pertaining to the interpretation of these data, the deficit and the developmental delay models. In the deficit model a causal relationship is generally implied between the antecedent event and the subsequent performance. The monumental body of literature devoted to the notion of "minimal brain dysfunction" (see Satz & Fletcher, 1980, for a critical review) attests to the widespread acceptance of this model. The developmental delay model, on the other hand, holds that the child will pass through all the stages of development but at a slower rate than the average child. The fact that the child is lagging behing the peer group at one point in development simply forecasts delays in the attainment of subsequent skills; a causal relationship is not inferred. The wide variety of different skill deficits which have been documented between learning disabled and normally achieving youngsters can largely be accounted for within this model. As this is a theoretical issue which has been intently debated elsewhere

(Fletcher & Satz, 1979; Rourke, 1976; Satz & Fletcher, 1980; Satz, Rardin, & Ross, 1971; Vellutino, 1979), it will not be further elaborated here. Suffice it to say that, although assessment procedures of young children are not necessarily affected by one's viewpoint in this regard, subsequent interpretation of the data and follow-up recommendations with regard to programming for the "at risk" group will differ radically depending upon the underlying theoretical framework.

A psychometric approach to prediction makes it possible to assess children at an early age before entrance into the school system. It would seem reasonable to expect an attenuation of the relationship between predictor and the criterion of academic performance depending upon how young the child was when first assessed. This section will examine the validity of using psychometric instruments with children between the ages of 4 and 6, both singly and in combination, as indices of future academic performance. Studies which employ teacher ratings or readiness tests in combination with psychometric measures will also be reviewed here.

A number of studies have attempted to predict future academic performance on the basis of measuring a single aspect of development such as sensory, motor, perceptual, linguistic, or cognitive functioning. The majority of these studies have been correlational in nature and have met with varying degrees of success. Shipp and Loudon (1964), in a study involving 115

grade one children, report a correlation of .51 between scores on the Draw a Man Test administered at the beginning of the year and total achievement as of the end of the year. They noted a sex difference in performance with girls scoring significantly higher than boys (girls $\bar{X}$ IQ=118.8; boys $\bar{X}$ IQ=107.8). In addition age and performance were significantly related for boys (r=.33). and essentially unrelated for girls (r=.06).

Keogh and Smith (1970), in their study involving 49 kindergarten children followed through to the end of grade five, found higher correlations between the Bender Gestalt Test and future academic performance for boys than for girls. Significant correlations were found for both sexes with reading achievement and for boys with math achievement but there was no correlation between the BGT results and future math performance for girls. Feshbach, Adelman, and Fuller (1974) report a correlation of .33 between the BGT administered in kindergarten and end of grade one reading performance (N=433). Ashmore (1980) reported a correlation of .33 between the BGT administered at the beginning of grade one and academic performance at the end of the year (N=49). Duffey, Ritter, and Fedner (1976) found similar results in their study involving a final sample of 82 children. They reported a correlation of .31 between both the Draw a Man Test and the Beery Test of Visual-Motor Integration and the end of grade two performance as measured by the total score on the Stanford Achievement Test. Flynn and Flynn (1978) also reported significant

correlations (.36 & .32) between the Draw a Man Test and the Test of Visual-Motor Integration administered at the beginning of the kindergarten year and end of grade two performance as measured by the California Achievement Test in their study involving a final sample of 81 children. Colarusso, Gill, Plankenhorn, and Brooks (1980) in a study involving 40 inner-city black boys reported only two correlations significant at the .05 level between a number of predictor variables (VMI, Motor-Free Visual Perception Test, Peabody Picture Vocabulary Test, the Auditory Memory, Auditory Reception, Verbal Expression and Grammatic Closure subtests of the Illinois Test of Psycho-linguistic Abilities) and performance on the Iowa Test of Basic Skills administered at the end of grade one. Only the Motor-Free Visual Perception Test and the Auditory Memory subtest of the ITPA emerged as significant predictors. The MVPT correlated .365 with the composite score of the ITBS and the AM subtest of the ITPA correlated .412 with the math score of the ITBS.

In an attempt to analyse the foregoing research, the question arises as to the practical utility of using any of the instruments cited as predictors of future academic performance. Although some of the correlations between certain aspects of development and future performance achieved statistical significance, they were generally quite low and accounted for a relatively small amount of the total variance. The research of Shipp and Loudon (1964) as well as that of Keogh and Smith (1970), however,

suggests that more significant relationships may have been obtained had the sex and age variables not been confounded.

Multiple correlation analyses have been utilized by a number of researchers in their effort to achieve a more predictive index of future academic achievement. Reynolds, Wright, and Wilkinson (1980), in a study involving 89 children, found moderate correlations with end of grade two performance using both a measure of language development (Test of Auditory Comprehension of Language) and a measure of perceptual-motor development (Beery Test of Visual-Motor Integration) administered just prior to kindergarten entry. The TACL correlated .54 and the VMI .53 with end of grade two performance as measured by the composite score of the SRA Achievement Series. Using both tests in multiple correlation yielded a multiple R of only .61. Keogh and Smith (1970) achieved higher and more consistent correlations with future academic performance for both boys and girls when they combined Bender Gestalt scores and kindergarten teacher ratings in a multiple correlational analysis. The multiple R with third and fifth grade reading scores was .67 and .69 respectively for girls and .61 and .71 for boys which represented a significant increase over using the Bender or the teacher rating in isolation.

The first really significant battery approach to the problem of screening for potential learning problems was undertaken by de Hirsch, Jansky, and Langford (1966). Their pilot

study was designed to identity children at the kindergarten level who were at risk of failing in reading by the end of grade two. Thirty-seven tests which sampled perceptuomotor organization, language development and readiness to cope with printed symbols were administered to 53 kindergarten children and later correlated with end of grade two reading results. From the nineteen significant correlations, the ten tests which best differentiated the failing from the passing readers were selected and constituted the Predictive Index. This index was later revised and modified by Jansky and de Hirsch (1972). They administered a new battery consisting of nineteen variables (ten from the Predictive Index plus nine new ones) to a much larger, more heterogeneous sample of kindergarten children (N=508) and followed it up two years later on the 355 children who were still available to be tested. Using a stepwise multiple regression technique a best predicting equation was developed. The final battery, named the Screening Index, consisted of five tests: Letter Naming, Picture Naming, Gates Word Matching, Bender Gestalt and the Binet Sentence Memory subtest. The multiple correlation coefficient between these tests and second grade reading was .66. This battery correctly identified at the kindergarten level roughly three out of four children who failed in reading (i.e. achieved a score of 2.2 or less on a silent paragraph reading test) at the end of grade two. To be exact (See Table 6), in the total sample of 355 children the

Table 6

Prediction of Grade Two Reading Performance using the Jansky and de Hirsch Screening Index.*

| | Silent Paragraph Reading Test | | |
|---|---|---|---|
| Screening Index | Poor (G.E. ≤ 2.2) | Adequate (G.E. > 2.2) | Total |
| Poor | 95 (79%) | 59 (25%) | 154 |
| Adequate | 25 (21%) | 176 (75%) | 201 |
| Total | 120 | 235 | 355 |

Total hits = 271/355 (76%) Total misses = 84/355 (24%)

*Data from Jansky and de Hirsch (1972).

Screening Index singled out 79% (95/120) of the failing readers and screened in 25% (59/235) of those children who eventually read at grade level. From the table it can also be seen, however, that the proportion of children identified by the test as "at risk" who actually experienced future problems was only .62. In other words, a child so labeled by the test still had about 2 chances in 5 of achieving adequately.

Feshbach et al. (1974) administered the original Predictive Index in the spring of the kindergarten year to two cohorts of children (888 in Year I and 844 in Year II of the study). Their follow-up of the first group two years later yielded less optimistic results than their report revealed. A reorganization of the data (See Table 7) indicated that a total hit rate of 73% (418/572) was indeed achieved but of the 170 children who later failed in reading (score of 1.7 or less on the Gates MacGinite) only 44 (26%) were identified by the PI using a cutoff score of $\leq 4$ as predictive of future learning problems. On the other hand, 93% (374/572) of the good readers were correctly identified.

Eaves, Kendall, and Crichton (1974), using a modified version of the PI and a cutoff score of $\leq 1$ as predictive of future problems, achieved somewhat more impressive results (See Table 8). Of the 30 children who were experiencing learning difficulties by the end of grade two 19 (63%) were correctly

Table 7

Prediction of Grade One Reading Performance on the Gates-MacGinite Reading Test using the de Hirsch Predictive Index.*

| Predictive Index | Gates-MacGinite Reading Test: Poor (≤1.7) | Adequate (>1.7) | Total |
|---|---|---|---|
| Poor | 44 (26%) | 28 (7%) | 72 |
| Adequate | 126 (74%) | 374 (93%) | 500 |
| | 170 | 402 | 572 |

Total hits = 418/572 (73%) Total misses = 154/572 (27%)

*Data adapted from Feshbach, Adelman, and Fuller (1974)

Table 8

Prediction of Grade Two Achievement as exemplified by Teacher Recommended Grade Placements using a Modified Predictive Index.*

| Modified Predictive Index | Recommended Grade Placement: Not ready for Grade 3 | Ready for Grade 3 | Total |
|---|---|---|---|
| Poor | 19 (63%) | 13 (10%) | 32 |
| Adequate | 11 (37%) | 120 (90%) | 131 |
| Total | 30 | 133 | 163 |

Total hits = 139/163 (85%) Total misses = 24/163 (15%)

*Data from Eaves, Kendall, and Crichton (1974).

identified. In addition, 129 of the 133 adequately achieving youngsters were also selected for a total hit rate of 85%.

A long-term predictive study which ran from 1970 to 1976 was undertaken by Paul Satz and his associates (Satz & Friel, 1972, 1974, 1978; Satz, Friel, & Rudegeair, 1976; Satz, Taylor, Friel, & Fletcher, 1978). This project was more strictly controlled and far more extensive in scope than any prior investigation. The standardization group consisted of the total population of white boys (N=497) who began kindergarten in the fall of 1970 in Alachua County, Florida. These children were followed to the end of grade five with criterion measures being taken at the end of grades one (1972), two (1973), three (1974), and five (1976). A determined effort to track these children resulted in a very low rate of attrition of about 10%. In addition, two cross-validation studies were completed, one on a smaller sample of white boys (N=181) who began kindergarten a year later, in 1971, and one on a mixed sample of kindergarten children (boys, girls, blacks, and whites, N=132) who were tested in 1974. Both of these studies were followed up at the end of the second grade.

A four-group discriminant function analysis was carried out for each of the above studies. The predictive classifications for those studies involving the standardization group were based on the kindergarten discriminant function composite scores. The predictive classifications for the cross-validation studies were based on the discriminant function weights of the standard-

ization population. Although the accuracy of the predictive classification of children into classroom reading level criterion groups varied somewhat over the series of analyses, a general statement can be made. Whether based on the total battery (19 variables), the abbreviated 8 variable battery or the experimental 5 variable battery (Finger-Localization, Recognition-Discrimination, Beery, Alphabet Recitation, Peabody Picture Vocabulary Test), the four-group discriminant function analyses consistently identified roughly 85-95% of the severe high risk and 85-95% of the superior readers in grades 2 through 5 with the overall hit rate being in the neighborhood of 72-75%. The accuracy of the battery was largely confined to the extreme groups (severe and superior) with the majority of prediction errors occurring in the mild and average groups. Using the end of grade two follow-up as an example (SeeTable 9) 89% of the severely delayed group and 94% of the superior group were correctly identified. Dichotomizing the data into two groups, "at risk" (severe and mild) verses "not at risk" (average and superior), in order to allow a more direct comparison with previously analyzed research reveals a valid positive rate of 79% (95/120) and a valid negative rate of 77% (260/338) for a total hit rate of 78% (355/458).

In addition to the predictive analyses, Satz and his associates also conducted stepwise multiple regression analyses in order to determine the discriminative ranking of the predictor

Table 9

Prediction of Grade Two Reading Level based on the Kindergarten Discriminant Function Composite Scores.*

| Composite Test Scores | Criterion Groups | | | | |
|---|---|---|---|---|---|
| | Severe | Mild | Average | Superior | Total |
| Poor (+) | 48 (89%) | 47 (71%) | 74 (27%) | 4 (6%) | 173 |
| Good (-) | 6 (11%) | 19 (29%) | 196 (73%) | 64 (94%) | 285 |
| Total | 54 | 66 | 270 | 68 | 458 |

Total hits = 355/458 (78%) Total misses = 103/458 (22%)

*Data from Satz, Friel, and Rudegeair (1976).

variables. For the standardization group the ranking remained essentially unchanged over time until the six year follow-up. In the first three studies Finger-Localization, Alphabet-Recitation and Recognition-Discrimination were the top ranking predictors. These instruments all loaded on Factor I, a general measure of sensory-perceptual-motor-mnemonic ability. Although the predictive ranking varied slightly by the end of grade 5 (Finger-Localization, Peabody, Beery, and Alphabet-Recitation), three of the top four variables again loaded on Factor I. The discriminitive ranking for the cross-validation studies, although maintaining Factor I tests in the top two positions, also deviated from the established pattern, likely as a function of sample selected.

An independent study using the abbreviated 8 variable test battery developed by Satz et al. was carried out by White, Batini, Satz and Friel (1979) on a sample of 407 white boys entering grade one in Perth, Australia and followed up three years later. As with the previous research, the tests correctly predicted 88% of the severe group and 94% of the superior group. Again the majority of prediction errors was confined to the mild and average groups. The overall prediction was 73% (242/332). The predictive ranking of the tests, however, was quite different from the American studies, likely reflecting the cultural and age differences in the test group (i.e. these children were almost one year older than the children in the American studies

when the predictor measures were administered and a high percentage of the children came from homes where English was not the main language). When word-recognition was the criterion measure, Factor I tests did predominate but when reading comprehension was the criterion measure, conceptual-verbal tests predominated.

Book (1974, 1980) employed a progressive screening procedure using measures of intelligence, school readiness, and perceptual-motor functioning in his attempt to identify educationally at risk children at the kindergarten level. In his first study (N=725) all children were administered the Metropolitan Readiness Test in April of their kindergarten year. Those who scored below the 25th percentile were then given the Slossen Intelligence Test. The Stanford-Binet was administered to those children who scored < 85 on the SIT. The Bender Gestalt was administered to all children who scored below the 25th percentile on the letter and form copying on the MRT. On the basis of the ensuing results, all children were assigned to one of six categories ranging from educable mentally handicapped to consideration for enrichment. A significant correlation was found between the end of second grade reading achievement and the diagnostic category to which the child was assigned. In his later study Book screened two cohorts of kindergarten children and followed them to the end of grade four. The Kindergarten Evaluation of Learning Potential (replacing the MRT) was administered to all kindergarten children in April as was the

Bender Gestalt. The SIT was given to those children who scored below the 25th percentile on the KELP and the Stanford-Binet was given to those children who scored < 85 on the SIT. Children who scored 75 or below on the S-B were eliminated from the study. The children were then placed in one of three groups: Group I comprised students who were below the 25th percentile on at least two of the measures, Group II comprised students who were below the 25th percentile on one of the measures, and Group III comprised students who were above the 25th percentile on all of the measures. Significant correlations were found between kindergarten group designation and subsequent academic performance in grades one through four (grade 1, r=.76; grade 2, r=.71; grade 3, r=.72; grade 4, r=.62). The results of a classificational analysis completed at the end of grade two using a group designation of I or II as indicative of "at risk" and III as indicative of "not at risk" and performance below the 50th percentile as indicative of learning difficulties revealed that this screening procedure correctly identified 82% of the population with a 12% false positive and a 23% false negative rate (Table 10).

It is evident that the numbers in the cells will change if the cutoff score of either the predictor or the criterion variable is altered. The data available from Book's research provides an example (Table 11). When the criterion cutoff score is lowered to the 25th percentile, a greater proportion of the more severely delayed group is identified by the screening procedure and the

Table 10

Prediction of Second Grade Reading Achievement Based on Kindergarten Group Designation.*

| Kindergarten Group Designation | Stanford Achievement Test: Poor (< 51%ile) | Adequate (≥ 51%ile) | Total |
|---|---|---|---|
| At Risk (I & II) | 76 (77%) | 11 (12%) | 87 |
| Not at Risk (III) | 23 (23%) | 83 (88%) | 106 |
| Total | 99 | 94 | 193 |

Total hits = 159/193 (82%) Total misses = 34/193 (18%)

*Data from Book (1974).

Table 11

Prediction of Second Grade Reading Achievement Based on Kindergarten Group Designation.*

| Kindergarten Group Designation | Stanford Achievement Test: Poor (< 26%ile) | Adequate (≥ 26%ile) | Total |
|---|---|---|---|
| At Risk (I & II) | 48 (92%) | 39 (28%) | 87 |
| Not at Risk (III) | 4 (8%) | 102 (72%) | 106 |
| Total | 52 | 141 | 193 |

Total hits = 150/193 (78%) Total misses = 43/193 (22%)

*Data from Book (1974).

false negative rate declines but at the same time many of the children labeled "at risk" do not manifest learning problems according to the new criterion and hence the false positive rate increases.

In attempting to analyse the research reviewed in this section and make some generalizations, a number of problems surface. In the first place, the incidence of learning problems as defined by the criterion variable in the majority of studies reviewed exceeded 20% of the population, a somewhat higher figure than might generally be expected except in cases of select groupings such as might be found in inner city schools or, as was the case in the research by Satz and his associates (1978), in an all male group. Such a high percentage suggests that our definition of learning problems might be too broad to be very useful. From a purely practical viewpoint it simply would not be feasible to introduce intervention programs with such a large proportion of the school population. If indeed 20% to 50% of the children are not succeeding in a given program, then it is likely the curriculum and/or the classroom environment rather than the children we should be looking at. A more useful approach would be to establish criterion cutoff scores which would identify various levels of functioning within the group. The efficiency of the predictor variables could then be judged in terms of the hit rate for a more specific group (i.e. severe verses mild deficit) as was possible in the research conducted

by Satz et al. referred to previously.

A second problem concerns the efficiency of the predictor variables and focuses on the philosophical issue as to whether to screen broadly or narrowly. In the first case, large numbers of children who, in fact, might never exhibit learning problems would be screened in initially and falsely labeled "at risk"; in the second instance, children who in fact might later develop problems would be missed by the screening procedure and falsely labeled "not at risk". The argument for keeping the number of false positives to a minimum focuses on the spectre of the self-fulfilling prophecy. Those who adhere to this belief argue that children falsely labeled will experience severe trauma as a result of that labeling and that low expectations on the part of parents and teachers will result in lower achievement. This stance begs the question as to whether or not remedial intervention should ever be offered if children are adversely affected when they are singled out. Furthermore, parents and teachers are given little credit for sensitivity, rudimentary intelligence and, in the case of teachers at least, extensive training in the skills necessary to meet the needs of the children entrusted to them.

Others argue that it is false negatives which should be kept to a minimum. They reason that children who are screened in but in reality have no difficulty (false positives) will quickly be identified on the basis of their successes once treat-

ment has begun. The key would be to have flexible groupings which could readily be altered as the need arose. False negatives, on the other hand, pose an entirely different problem. If the screening device missed large numbers of children who actually needed assistance, they would not likely be noticed until much later after they had repeatedly failed at tasks their peers had handled successfully. It is precisely this situation we are attempting to circumvent by the use of early screening procedures.

A possible solution to the dilemna of whether or not to offer treatment in individual cases may be found in the decisional procedure proposed by Satz et al. (1976). In this study the test signs + and - were subdivided to generate four levels of test decisions: severe high risk (++), high risk (+), low risk (-), and very low risk (--). The conditional probabilities were then computed for each of these signs. It was found that a decision to initiate treatment of the basis of a severe high risk sign (++) would be correct in 82% of the cases, whereas initiating treatment given a less high risk sign (+) would be correct only 44% of the time. In other words, offering treatment to children at this level would result in more children who didn't need treatment receiving it than children who did. This type of detection procedure would allow flexibility in decision making based on such factors as the general incidence of severe high risk children in the school, the relative

risks associated with intervention verses non-intervention and the availability of resources.

A third problem in drawing conclusions from the research centers on the variability among the predictor and criterion variables themselves. Not only are different types of instruments used ranging from teacher judgments to a number of standardized instruments but the scores themselves may reflect national, local, school or even classroom norms. Furthermore, different types of scores are used ranging from raw scores, z-scores, and percentile ranks to age- and grade-equivalents. It is evident that certain types of scores are more useful than others. Berk (1981) argues convincingly that the age- or grade-equivalent score should never be used. In the first place it is not an equal interval scale and as such arithmetic calculations cannot be performed. More importantly, however, although significance is generally attached to aspecific grade-equivalent there is no standardized meaning associated with any particular score. A score of 2.3 on a reading test and a score of 2.6 on a math test implies a difference in ability but, in fact, it is probable that no real difference exists. Only scores which indicate a child's relative position in the group (i.e. percentiles, z-scores, T-scores) can be compared with any validity.

The final difficulty in drawing conclusions from the research concerns the lack of comparability in the prediction process itself including both the instruments and the procedures used. The

procedures include correlational, multi-correlational, simple classificational and discriminant function analyses and the variables range from single instruments to multiple battery approaches. However, in spite of this broad spectrum, certain instruments seem to emerge as consistent predictors of future performance, namely those which assess aspects of sensory, motor, perceptual, and cross-modal functioning. In the SEARCH battery developed by Silver (reported in Satz et al., 1979) tests of visual perception (including discrimination, recall, and visual-motor control) contributed significantly in forecasting reading and learning problems. The top ranking predictors in the battery employed in the research conducted by Satz and his associates consistently reflected sensory, motor, perceptual, and mnemonic abilities as did the four top ranking tests in the de Hirsch battery as well.

In terms of the present study, it is predicted that the paper and pencil tests administered to the kindergarten children, in that they measure skills which research has shown to be effective forecasters of achievement when used with an appropriate age group, will be significantly correlated with the criterion variables. Furthermore, it is predicted that these instruments will be effective in identifying the "at risk" versus the "not at risk" groups in the population under consideration.

# CHAPTER 3

## Design and Procedures

This chapter will examine the technical aspects of the present research study. Included here will be a description of the subjects, including how they were selected, followed by an outline of the procedural and temporal framework within which the research was conducted. The assessment instruments and the rationale for their inclusion in the study will be discussed in the third section and the final section will focus on the statistical procedures used in the treatment of the data.

### Subjects

The subjects comprised all of the kindergarten children who were registered in the English language program in four schools in the County of Strathcona and who had been in attendance for a least two months prior to the beginning of the study. In all there were eight classroom groupings and a total of 205 children of whom 106 were boys and 99 were girls. Of these children one was eliminated at the kindergarten level due to incomplete data, 22 transferred out of the area before final data collection late in the grade one year could be completed, and 10 were retained in kindergarten leaving a final sample of 172 (91 boys and 81 girls) for whom complete information was available. This represented an attrition rate of

roughly 16%.

The four schools which took part in this project, Brentwood, Campbelltown, Mills Haven and Glen Allan are all located within the suburban center of Sherwood Park and were selected on the basis of their expressed willingness to participate. These schools are representative of schools in the suburban areas of the County of Strathcona and are likely comparable to schools in other suburban communities as well. Collectively they service low-middle to upper-middle class families with occupations ranging from the blue collar and service industries to the white collar and professional classes.

## Procedure

Initially, a letter describing the intent and proposed design of the project was sent out to the principals of all the public schools located within the boundaries of Sherwood Park (A copy can be found in Appendix A). Because of the anticipated difficulties excessive distances would have presented to the researcher, the rural schools and those located within the town of Fort Saskatchewan were not included. A follow-up meeting with the principal, or principal-delegate, as well as affected staff (counsellors, teachers and aides) was scheduled in each of the schools which had expressed an interest in the project to discuss what would be required in terms of staff and student involvement and answer any questions

which might be forthcoming. Once the schools had agreed to participate, letters on individual school letterhead were sent home to inform the parents of the project and to allow them time to react to it (See Appendix B). There were no negative reactions and no children were excluded from the study at this stage.

The study was designed in such a way as to only minimally disrupt routine classroom procedures. The guiding philosophy here was that if a testing procedure was to be useful on a large scale basis, responsibility for administration and scoring must, potentially at least, lie with the classroom teacher. With this in mind, the test administration procedures were adapted somewhat to comply with the instructional methods and physical environment of the typical kindergarten classroom. The teacher explained to the large group, held at the beginning of each class, that a particular drawing task was to be carried out at a specific center under the supervision of either herself or her aide. At this time only general instructions designed to familiarize the children with the task were given. More specific instructions were given once the children were at the center in question. The Group Perceptual Screening Test (GPST) and the Beery Test of Visual-Motor Integration (VMI) were administered to small groups of between four and seven children whereas the Draw a Person Test (DAP) was given to only two children at a time. In this manner the entire test battery

was administered over a two week period at the beginning of April, 1982. It was considered that such a time span would be flexible enough to allow for scheduled field trips and other activities which occur as a regular part of any kindergarten program and yet restrictive enough such that no differences attributable to increased maturity would occur.

The tests were administered in a specific order. The DAP, in that it is generally considered to be an enjoyable task for this age group of children and at the same time an easy task to explain, was given first followed by the VMI. The GPST was scheduled last. It was felt that since the children were asked to formulate concrete verbal associations with each of the designs in this test the procedure might carry over to the VMI and perhaps contaminate the results.

During the same time period the kindergarten teachers were asked to complete a modified version of the Pupil Rating Scale (PRS) on each of their students and to rate the child's general readiness for grade one in terms of "not at all ready", "average readiness" and "highly ready". This three point rating scale was included in an attempt to clarify the issue as to whether or not a rather lengthy, detailed rating scale would yield a more accurate assessment of the child than a very simple scale.

In addition to these instruments, the teachers also completed an information sheet on each child which made possible

the identification of a number of specific groups of children about whom school personnel and parents alike have expressed a variety of concerns. It seemed feasible to identify these groups and, if the numbers were sufficient, examine their performance on the various instruments to determine whether they differed significantly from the rest of the group. Specifically these groups were: children on Category A grants, children in daycare (CARE) and children repeating kindergarten.

All test protocols were scored by the researcher. Although the scoring of these instruments is certainly within the capabilities of the classroom teacher, a review of the literature revealed that the interscorer reliability, especially for the VMI (Pryzwansky, 1977), may be quite low with untrained personnel. Due to the time constraints, training of the teachers in these procedures did not seem feasible and hence the responsibility for scoring the kindergarten tests remained with the experimenter.

The Metropolitan Readiness Test (MRT) was included in this study as it is given routinely to all incoming grade one pupils in the County and hence the scores were readily available. It was administered late in September and hand-scored by the individual classroom reachers. The final follow-up was carried out at the beginning of March, 1983. The grade one teachers were asked to rank those students who were in the study in terms of their general functioning in each of three areas: reading, language and arithmetic skills (See Appendix A). As previous research

indicated that prediction was most accurate for the extreme groups, the ranking scale was specifically designed to identify those children who fell at the extremes of the total distribution. The teachers were asked to reflect upon each student and decide whether he or she would be considered to be average, below average, or above average in each area. As a second step they were asked to look at those students whom they had designated as either below or above average and decide whether any child in those groups might be considered as "clearly inferior" or "clearly superior" to the others. A caution was given to the effect that, except in rare instances, no more than one to three children in a class of 30 would be expected to fall into either of these two groups. This procedure was followed for each of the three academic areas and the scores were then summed to yield a general academic index. It was felt that having the teachers rank the children in three separate areas would yield a more reliable index of their overall level of academic functioning. Only the total score was used in the analyses.

Instrumentation

The kindergarten tests administered in this study were chosen for a number of reasons. In the first place two of the instruments, the VMI and the DAP are used almost routinely by psychologists, school counsellors and health unit personnel in their assessment of young children. One of the aims of

this study was to assess the validity of using these instruments as screening devices for academic purposes. The GPST was included because of its similarity to the VMI (i.e. both are form-copying tests) and because, to the best of this researcher's knowledge, it has not been widely used and hence remains an unknown quantity. Likewise, the PRS has not often been used with kindergarten populations. This study will examine the utility of including these two instruments in a battery approach to the prediction of grade one academic performance. Secondly, the use of drawing tasks is easily understood by and generally highly motivating for young children and hence circumvents, partially at least, the difficulty in obtaining with some degree of consistency a young child's best effort. However, in interpreting any test results which involve young children, it is important to remain cognizant of the warning advanced by Keogh and Smith (1970) that "findings of deficiency...may be less valid for school prediction than is specification of competencies" (289). The final reason why only drawing tasks were used in the early assessment phase of the study relates to a substantial body of literature which conceptualizes psychological and neurological development as progressing sequentially and in hierarchical stages proceeding from a state of diffuse organization to greater differentiation and integration. Since this maturational process is essentially age-linked, assessing a child's maturational level and comparing

this to chronological age will yield information as to developmental rate and hence, theoretically at least, allow predictions as to future development. It is important, however, to choose tests which are both maturation sensitive and age-appropriate (i.e. assess those skills which are undergoing primary development during a specific age span). The three tests chosen for this study (VMI, DAP, GPST) assess aspects of visuomotor functioning including visual perception of and alertness to detail, visual memory, visual-motor integration and fine motor control, all skills considered to be in primary ascendancy between the ages of four and seven. The theory predicts that a maturational lag in perceptual-motor development forecasts a similar delay in conceptual-cognitive development in that the former underlies and subsumes the latter (Satz & Sparrow, 1970; Satz & van Nostrand, 1972). In terms of this study it is expected that those children who manifest a maturational lag as measured by the VMI, the GPST, and the DAP will not be ready for the higher order cognitive tasks of the first grade classroom.

Draw a Person Test (DAP). Because of its simplicity, inherent motivational properties and ease of administration, the human figure drawing is likely one of the most commonly used instruments in the assessment of young children. Interpretations of such drawings range from attempts at diagnosing personality and emotional disorders to the assessment of intellectual

functioning. In this study the task is considered to reflect the child's level of perceptual-motor functioning and hence the conceptual-cognitive development as well, insofar as the former is considered to underlie and be subsumed by the latter in terms of the theoretical bias of this study.

Administration procedures are simple and straightforward. The child is given a standard-sized piece of plain paper and a pencil and asked to draw a picture of a whole person. Scoring the protocols, on the other hand, is tedious and time consuming. The system developed by Harris (1963) was utilized in that it provided separate sets of well-illustrated and detailed scoring criteria for drawings of both men and women. In addition, raw scores could then be converted to standard scores having a mean of 100 and a standard deviation of 15. These norms were established on a sample of almost 3,000 children between the ages of 5 and 15.

Test retest reliabilities for the DAM test are reported in the .70's to .80's and interscorer reliabilities reportedly range from .91 to .98. In addition to the face validity of this instrument which cites increasing mean scores with increasing age, the DAP has been extensively studied in terms of its correlation with measures of achievement and intelligence. Shipp and Loudon (1964) report a correlation of .51 between DAM scores and end of grade one achievement. Moderate correlations ranging from .36 to .65 are reported by Harris (1963) between DAM scores and Stanford-Binet IQs. Ritter, Duffey, and

Fischman (1974), using a kindergarten sample, found a correlation of .55 between these two instruments but noted that the DAM tended to underestimate IQ and that this tendency became more marked as IQ increased, a phenomenom also noted by Reisman and Yomokoski (1973).

The Developmental Test of Visual-Motor Integration (Beery, 1967). The VMI is a form-copying test consisting of 24 designs arranged sequentially by level of difficulty. It is purported to measure aspects of a child's visual-motor functioning including visual perception, visual-motor integration, and fine motor control. The short form consisting of 18 designs is appropriate for use with children between the ages of 2 and 8 and can be individually or group administered (Pryzwansky, 1977). The children are simply asked to copy the forms without tilting the booklet, erasing or reworking the designs. In this study 5 to 7 children were given the test at one time with either the teacher or the aide acting as the monitor. The children did not turn the page to the next set of designs until instructed to do so. In this way it was possible to assist the children with turning the pages so that the proper sequence was maintained and to ensure that the booklets were correctly oriented. The scoring system is relatively straightforward. Each design is simply checked against a set of well illustrated criteria and scored as either "pass" or "fail". The raw score is determined by tallying the number of successes achieved before three successive failures occur.

Age-equivalent norms based on the total raw scores are presented separately by sex. Although high interscorer reliability has been noted (Ryckman & Rentfrow, 1971; Snyder & Snyder, 1981), Pryzwansky (1977) cautions against using untrained personnel. In this study all protocols were scored by the researcher.

The VMI is generally considered to possess sufficient reliability to be used with young children (Ryckman & Rentfrow, 1971) and it has been found to correlate significantly with future academic performance (Duffey, Ritter, & Fedner, 1976). Perhaps the most significant type of validity information in terms of the present study is that pertaining to the VMI's high correlation with chronological age specifically in the 5 to 7 year old age range (Duffey et al., 1976; Harris, 1967; Porter & Binder, 1981) indicating that it is a maturation sensitive instrument.

The Group Perceptual Screening Test (Kelly, 1970). The GPST is a short form-copying test intended for use with preschool and first grade children as a means of identifying children with deficient perceptual-motor functioning. The test consists of five line drawings all between two and three inches in size presented on two standard sized sheets of plain paper. Three are placed on the left hand side of the first sheet and the remaining two are on the second sheet. Administration procedures differ substantially from those of other form-copying tests. The teacher directs the children's attention to the

first drawing and asks them to associate it with something from their own experience. After receiving several responses the teacher demonstrates on the chalkboard how to execute the design. The children are then directed to copy the one on their sheet in the appropriate space. All five figures are completed in this manner. There is an objective scoring system with criteria carefully delineated for each of the three points which are allotted for each figure. The total possible score is 15. Inter-scorer reliability on a 300 student first grade sample was reported by the test author to be .94. Norms, although not extensive, were established on a sizeable group of children at each of three different levels of schooling: first semester kindergarten (N=632, $\bar{X}$ age=5-6), spring of the kindergarten year (N=2,165, $\bar{X}$ age=6-0), and early grade one (N=3,728, $\bar{X}$ age=6-6). Means and standard deviations as well as critical cutoff scores for three levels of deficiency ("borderline", "low", and "very low") are given for each of these groups. To the best of this researcher's knowledge the validity of this instrument has not been adequately established. The test author cites only one study. In a sample of children identified as "at risk" by the GPST, 90% were also found to exhibit deficiencies in one or more areas as measured by the Frostig (Kelly, 1970).

The Pupil Rating Scale (Myklebust, 1971). The PRS was originally developed as a screening instrument for learning disabilities to be used by the classroom teacher. It consists

of 24 items grouped into five subtests; auditory comprehension and memory, spoken language, orientation, motor coordination and personal-social behavior. The items within each of the subtests comprise specific observable behaviors to be rated by the classroom teacher on a scale from 1 (poorest performance) to 5 (best performance). The ratings themselves are clearly defined in behavioral terms for each individual item.

The PRS was originally normed on 2,176 third and fourth grade pupils. The means and standard deviations for the five subtests, two subtotals and grand total are presented by grade (third and fourth), sex and group (learning disabled, borderline and normal control).

The validity information available consists mainly of inter-group comparisons. It was found that all eight scores significantly differentiated between groups. In a replication study involving 516 third and fourth grade students, Pihl and Nagy (1980) found no significant differences in the means and standard deviations from those reported in the standardization sample. Colligan (1977, 1979), using a group of 60 kindergarten children found significant correlations between the PRS and several criterion measures administered both concurrently and two years later. In terms of predictive validity, similar to Myklebust's work involving third and fourth grade pupils, he found a systematic progression of worsening scores with the degree of learning difficulty. (The children were

categorized as "no disability", "mild disability", "moderate disability", and "severe disability" according to the amount of remedial assistance provided over their three years of schooling.)

For the purposes of this study the PRS was shortened.[1] Only those subtests which measured behavior easily observable by the kindergarten teacher and not previously tapped by the DAP, VMI, or GPST were retained. In this way the length of the scale was reduced from five to three subtests (auditory comprehension and memory, spoken language, and personal-social behavior) each consisting of 12 items.

The Metropolitan Readiness Tests (Nurss & McGauvran, 1976)

The MRT were originally developed for the assessment of a child's readiness to begin formal reading instruction. The first edition appeared in 1933 and was subsequently revised in 1949, 1964, and 1976. There are two levels of the 1976 edition: Level I designed for use from the beginning through the middle of kindergarten and Level II for use at the end of kindergarten and the beginning of grade one.

Level II is composed of eight subtests which purport to measure basic skills needed in beginning reading and arithmetic and yields scores in four separate skill areas plus a prereading composite score. For the purposes of this study, however, only

1. Stevenson et al (1976) found that a short rating scale was as efficient as a long one for prediction of contemporary or subsequent achievement.

the Prereading Skills Composite score will be used as the technical information available in the manual indicates that this score possesses greater reliability and predictive validity than any of the individual subtest scores. Alternate form reliability is listed as .89 and split-half reliability as .94 for this score. Predictive validity with the Metropolitan Achievement Tests Total Reading score as the criterion was .70 (N=1,997). When the Stanford Achievement Test Total Battery Score was used as the criterion, the correlation was .76 (N=2,024).

The MRT was considered for inclusion in this research study mainly because it is administered routinely to grade one classes in the County of Strathcona and hence the scores were readily available. On the other hand, it is a widely used screening instrument (Maitland et al., 1974) and, although a great deal of research has been done on the relationship between performance on the MRT and future achievement, little seems to be known about the relationship between kindergarten teachers' perceptions of a child's potential and subsequent performance on the MRT. A unique opportunity existed not only to shed light on this relationship but also to discover whether or not the MRT contributed significantly in the correct classification of pupils or whether it might, in fact, be superfluous.

Data Analyses

The data collected for this research project were first subjected to a descriptive analysis. The means, standard deviations and intercorrelations for all variables were computed for the total group and separately by sex. The resulting sex differences were tested for significance by means of a multivariate technique. The data were then analyzed for the group as a whole and separately by sex using a stepwise multiple regression procedure to determine whether using the variables in multiple correlation would significantly add to the amount of explained variance or whether they were in fact redundant. Finally, classificational analyses were undertaken to clarify where prediction errors would occur given a specific set of criteria.

# CHAPTER 4

## Results

This chapter examines the results of the statistical procedures carried out during the course of the study. The first section deals with the descriptive statistics and focuses on a discussion of the means, standard deviations, and intercorrelations of the variables presented for the group as a whole and separately by sex. The means and standard deviations are presented in Table 12; the correlation matrices in Tables 14, 15 and 16. The equivalency of means across sexes was tested and significant differences were found. These results are shown in Table 13. A similar investigation was also undertaken to determine whether there were significant differences in means between that group of students which moved out of the system and the group which remained. As no significant differences were found the results of this analysis may be found in Appendix C. The second section of this chapter focuses on a discussion of the results of the multiple regression analyses which are summarized in Tables 17 and 18. Finally, in keeping with that growing body of research which assesses the effectiveness of prediction procedures by means of classificational analyses, the last section of this chapter looks at the data from this perspective.

Descriptive Statistics

The means and standard deviations for the variables employed in this research study are found in Table 12, presented for the total group and separately by sex. A cursory inspection of this data reveals the presence of sex differences on almost every variable. The very similar scores on the "CARE" variable indicates that the number of boys and girls in each of the two conditions (i.e. 1=homecare; 2=daycare) was approximately the same with the majority in "homecare". Univariate tests of significance on the remaining variables indicated that real differences existed in only eight of the comparisons (See Table 13). In the first place girls are significantly younger than boys. This likely results from the fact that more boys than girls seem to be retained in kindergarten. On the 16 children who were in their second year of kindergarten 13 were boys and 3 were girls; of the 10 children who did not proceed to grade one in the fall of 1982, 7 were boys and 3 were girls. Secondly, it is interesting to note that on two of the three drawing tasks administered at the kindergarten level (GPST, DAP) girls performed significantly better than boys whereas on the third (VMI), there was no significant difference between the sexes. Predictably, as evidenced by the scores on tht PRS as well as on the simple rating scale, teachers found girls significantly more ready to enter grade one than boys. Subtest 3 of the PRS, personal-social behavior, would seem to be the most influential factor in this boy-girl discrepancy although scores on the other two subtests were slightly higher for girls as well.

Table 12

Means and Standard Deviations of all Variables

| Variable | Boys | | Girls | | Total Group | |
|---|---|---|---|---|---|---|
| | $\bar{X}$ | SD | $\bar{X}$ | SD | $\bar{X}$ | SD |
| Age (in months) | 71.15 | 4.66 | 69.74 | 3.30 | 70.49 | 4.06 |
| Care | 1.21 | 0.41 | 1.16 | 0.37 | 1.21 | 0.41 |
| GPST Raw Score | 8.81 | 2.22 | 9.59 | 2.29 | 9.18 | 2.37 |
| DAP Raw Score | 13.98 | 4.81 | 17.83 | 5.45 | 15.79 | 5.49 |
| DAP IQ | 91.48 | 12.94 | 95.77 | 13.13 | 93.50 | 13.10 |
| VMI Raw Score | 10,10 | 1.85 | 10.46 | 1.81 | 10.27 | 1.93 |
| VMI Age | 68.52 | 8.39 | 68.33 | 6.97 | 68.43 | 7.98 |
| VMI Discrepancy | 22.36 | 8.96 | 23.69 | 6.98 | 22.99 | 8.21 |
| PRS Total | 37.21 | 6.03 | 39.91 | 8.29 | 38.48 | 7.78 |
| PRS Sub. 1 Aud. Comp. | 12.67 | 2.67 | 13.28 | 3.09 | 12.96 | 3.08 |
| PRS Sub. 2 Spoken Lang. | 12.58 | 2.36 | 13.35 | 2.96 | 12.94 | 2.81 |
| PRS Sub. 3 Pers.-Social | 11.95 | 2.42 | 13.28 | 2.74 | 12.58 | 2.71 |
| Rating | 2.07 | 0.52 | 2.32 | 0.57 | 2.19 | 0.61 |
| MRT Raw Score | 53.35 | 12.43 | 56.33 | 10.26 | 54.76 | 11.80 |
| MRT Stanine | 5.52 | 1.78 | 5.88 | 1.62 | 5.69 | 1.73 |
| Grade 1 Academic Standing | 9.25 | 2.45 | 10.09 | 2.82 | 9.65 | 2.66 |

Note Boys N=91 Girls N=81 Total N=172

Table 13

Differences in Means (Boys versus Girls)

| Variable | Boys | Girls | F | P Level |
|---|---|---|---|---|
| Age | 71.15 | 69.74 | 5.134 | .025* |
| GPST | 8.81 | 9.59 | 5.044 | .026* |
| DAP Raw Score | 13.98 | 17.83 | 24.579 | .000* |
| DAP IQ | 91.48 | 95.77 | 4.688 | .032* |
| VMI Raw Score | 10.01 | 10.46 | 1.586 | .210 |
| VMI Age | 68.52 | 68.33 | .023 | .879 |
| VMI Discrepancy | 22.36 | 23.69 | 1.108 | .294 |
| PRS Total | 37.21 | 39.91 | 5.990 | .015* |
| PRS Sub. 1 Aud. Comp. | 12.67 | 13.28 | 1.923 | .167 |
| PRS Sub. 2 Spoken Lang. | 12.58 | 13.35 | 3.518 | .062 |
| PRS Sub. 3 Pers.-Social | 11.95 | 13.28 | 11.607 | .001* |
| Rating | 2.07 | 2.32 | 8.899 | .003* |
| MRT Raw Score | 53.35 | 56.33 | 2.920 | .089 |
| MRT Stanine | 5.52 | 5.88 | 1.930 | .167 |
| Grade 1 Academic Standing | 9.25 | 10.09 | 4.278 | .040* |

*Denotes significance at .05 level or better.

The differences on the MRT did not achieve significance at the .05 level whereas the general grade one academic standing as measured near the end of grade one once again reflected the male-female discrepancy, albeit at a level which just achieved significance.

The correlation matrices which are found in Tables 14, 15, and 16 also reflect certain sex differences. To be significant at the .05 level for the total group, correlations must be at least .13. In order to achieve this level of significance for the subgroups (i.e. boys, girls) these correlations must be .18 or greater. In terms of this level of significance, all of the variables except "age" are correlated significantly with both the MRT and grade one general academic standing. From the tables (15 and 16) there would seem to be a stronger relationship between certain variables (VMI, Subtest 3 of the PRS, the simple rating scale, and the MRT) and the criterion (grade one academic standing) for girls than for boys. The GPST, the DAP, Subtests 1 and 2 of the PRS and the PRS total, on the other hand, yielded similar correlations with grade one academic standing for both groups.

Multiple Regression Analyses

Multiple regression analyses were undertaken to ascertain, first of all, whether combining a variety of measurement devices (teacher evaluations, psychometric devices, and readiness tests) would, in fact, improve upon prediction and, secondly, to determine which combination of independent variables would best predict

Table 14

Correlation Matrix for All Variables (Total Group)*

| | Sex | Age | Care | GPST | DAPR | DAPI | VMIR | VMIA | VMID | PRST | PRS1 | PRS2 | PRS3 | Rat. | MRTR | MRTS | GR1 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Sex | - | -17 | -07 | 17 | 35 | 16 | 10 | -01 | 09 | 19 | 12 | 15 | 25 | 24 | 13 | 11 | 16 |
| Age | | - | -11 | 21 | 08 | -08 | 15 | 18 | -34 | 13 | 19 | 13 | 04 | 05 | 03 | 08 | 07 |
| Care | | | - | -01 | -15 | -12 | 03 | 03 | 08 | -10 | -07 | -05 | -14 | -12 | 02 | 03 | -03 |
| GPST | | | | - | 41 | 34 | 53 | 50 | 37 | 38 | 35 | 30 | 35 | 31 | 30 | 29 | 37 |
| DAP RS | | | | | - | 94 | 38 | 36 | 32 | 41 | 36 | 38 | 36 | 37 | 33 | 33 | 38 |
| DAP IQ | | | | | | - | 33 | 33 | 37 | 34 | 28 | 31 | 31 | 30 | 30 | 29 | 31 |
| VMI RS | | | | | | | - | 98 | 86 | 44 | 46 | 37 | 35 | 39 | 45 | 44 | 50 |
| VMI Age | | | | | | | | - | 86 | 43 | 46 | 36 | 33 | 35 | 42 | 42 | 48 |
| VMI Discrepancy | | | | | | | | | - | 34 | 34 | 28 | 30 | 31 | 39 | 36 | 43 |
| PRS Total | | | | | | | | | | - | 94 | 93 | 80 | 74 | 43 | 47 | 56 |
| PRS 1 | | | | | | | | | | | - | 90 | 59 | 64 | 39 | 42 | 54 |
| PRS 2 | | | | | | | | | | | | - | 57 | 65 | 41 | 44 | 54 |
| PRS 3 | | | | | | | | | | | | | - | 68 | 34 | 40 | 42 |
| Rating | | | | | | | | | | | | | | - | 33 | 33 | 48 |
| MRT RS | | | | | | | | | | | | | | | - | 95 | 60 |
| MRT Stanine | | | | | | | | | | | | | | | | - | 60 |
| Grade 1 Academic Standing | | | | | | | | | | | | | | | | | - |

*Decimals have been omitted.
Correlations equal to or greater than .13 are significant at $P \leq .05$.

Table 15

Correlation Matrix for All Variables (Girls)*

| | Age | Care | GPST | DAPR | DAPI | VMIR | VMIA | VMID | PRST | PRS1 | PRS2 | PRS3 | Rat. | MRTR | MRTS | Grl |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Age | - | 10 | 22 | 14 | -09 | 23 | 24 | -23 | 17 | 26 | 13 | 08 | 14 | 03 | 11 | 12 |
| Care | | - | 08 | -21 | -21 | -04 | -04 | -09 | -22 | -23 | -23 | -16 | -20 | -25 | -18 | -13 |
| GPST | | | - | 35 | 29 | 52 | 50 | 39 | 30 | 32 | 29 | 24 | 28 | 32 | 32 | 37 |
| DAP RS | | | | - | 94 | 29 | 29 | 27 | 33 | 33 | 34 | 27 | 35 | 33 | 34 | 38 |
| DAP IQ | | | | | - | 22 | 22 | 30 | 28 | 26 | 29 | 25 | 29 | 33 | 31 | 32 |
| VMI RS | | | | | | - | 99 | 88 | 40 | 43 | 37 | 34 | 49 | 49 | 47 | 56 |
| VMI Age | | | | | | | - | 88 | 40 | 42 | 36 | 34 | 48 | 48 | 46 | 56 |
| VMI Discrepancy | | | | | | | | - | 31 | 29 | 30 | 30 | 40 | 47 | 41 | 51 |
| PRS Total | | | | | | | | | - | 97 | 96 | 90 | 78 | 47 | 52 | 56 |
| PRS 1 | | | | | | | | | | - | 93 | 80 | 78 | 44 | 48 | 54 |
| PRS 2 | | | | | | | | | | | - | 77 | 79 | 42 | 49 | 55 |
| PRS 3 | | | | | | | | | | | | - | 65 | 45 | 49 | 50 |
| Rating | | | | | | | | | | | | | - | 39 | 38 | 54 |
| MRT RS | | | | | | | | | | | | | | - | 94 | 65 |
| MRT Stanine | | | | | | | | | | | | | | | - | 67 |
| Grade 1 Academic Standing | | | | | | | | | | | | | | | | - |

*Decimals have been omitted.

Correlations equal to or greater than .18 are significant at $P \leq .05$.

Table 16

Correlation Matrix for All Variables (Boys)*

| | Age | Care | GPST | DAPR | DAPI | VMIR | VMIA | VMID | PRST | PRS1 | PRS2 | PRS3 | Rat. | MRTR | MRTS | Grl |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Age | - | -25 | 27 | 18 | -04 | 13 | 15 | -38 | 20 | 19 | 19 | 10 | 07 | 07 | 10 | 09 |
| Care | | - | -06 | -06 | -04 | 10 | 07 | 20 | 05 | 09 | 13 | -11 | -02 | 20 | 20 | 09 |
| GPST | | | - | 41 | 36 | 52 | 52 | 34 | 43 | 37 | 28 | 41 | 29 | 25 | 23 | 33 |
| DAP RS | | | | - | 96 | 45 | 47 | 35 | 43 | 35 | 37 | 33 | 27 | 28 | 29 | 32 |
| DAP IQ | | | | | - | 41 | 42 | 41 | 37 | 28 | 31 | 32 | 26 | 26 | 26 | 27 |
| VMI RS | | | | | | - | 98 | 85 | 49 | 48 | 35 | 34 | 28 | 41 | 42 | 43 |
| VMI Age | | | | | | | - | 86 | 52 | 51 | 38 | 36 | 28 | 39 | 38 | 43 |
| VMI Discrepancy | | | | | | | | - | 38 | 38 | 26 | 28 | 22 | 33 | 32 | 36 |
| PRS Total | | | | | | | | | - | 90 | 88 | 65 | 65 | 40 | 42 | 54 |
| PRS 1 | | | | | | | | | | - | 86 | 32 | 46 | 34 | 36 | 52 |
| PRS 2 | | | | | | | | | | | - | 29 | 46 | 39 | 38 | 51 |
| PRS 3 | | | | | | | | | | | | - | 67 | 23 | 28 | 28 |
| Rating | | | | | | | | | | | | | - | 26 | 26 | 38 |
| MRT RS | | | | | | | | | | | | | | - | 95 | 56 |
| MRT Stanine | | | | | | | | | | | | | | | - | 53 |
| Grade 1 Academic Standing | | | | | | | | | | | | | | | | - |

*Decimals have been omitted

Correlations equal to or greater than .18 are significant at $P \leq .05$.

a) performance on the MRT administered in the fall of the grade one year and b) end of grade one general academic standing. In order to avoid spurious results only those variables which correlated significantly with the criterion and, at the same time, had relatively low intercorrelations were used in the regression analyses. Table 17 summarizes the regression analyses for the total group and separately by sex using the MRT as the criterion variable; Table 18 uses grade one general academic standing as the criterion and includes the MRT as a predictor. However, as a sub-objective of this study was to determine the effectiveness of using only information gathered during the kindergarten year as predictors of future achievement, the MRT data was programmed to enter the equation as a final step.

Table 17 focuses on the prediction of the MRT results. When the group is considered as a whole, it is evident that only two variables, the VMI and the PRS, contribute significantly to the regression equation. Although using these two variables in combination does increase the multiple correlation coefficient substantially, the end result remains unimpressive with more than 70% of the variance left unexplained. When the regression was run separately for boys and girls the demographic variable, CARE,[1] contributed significantly to the prediction equation but, once again, the total variance accounted for was less than satisfactory. Interestingly, more variance was explained by the testing procedures for girls than for boys (37.42% as opposed to 25.95%).

1. The other two categories, "repeaters" and "Category A" grant students were removed from the analysis due to insufficient numbers.

Table 17

Multiple Regression with Metropolitan Readiness Test (MRT) Total as Criterion

| Step | Variable | F Value | Mult.R | % Variance | % Variance Added | Simple r |
|---|---|---|---|---|---|---|
| **Total Group** | | | | | | |
| 1 | VMI Raw | 14.83* | .458 | 20.95 | - | .458 |
| 2 | PRS Tot. | 11.93* | .527 | 27.73 | 6.78 | .442 |
| 3 | DAP Raw | 2.40 | .535 | 28.66 | .93 | .328 |
| 4 | CARE | .48 | .537 | 28.86 | .20 | -.009 |
| **Boys** | | | | | | |
| 1 | VMI Raw | 4.84* | .421 | 17.74 | - | .421 |
| 2 | PRS Tot. | 4.76* | .478 | 22.83 | 5.09 | .411 |
| 3 | CARE | 3.29* | .504 | 25.45 | 2.62 | .175 |
| 4 | DAP Raw | .59 | .509 | 25.95 | .50 | .289 |
| **Girls** | | | | | | |
| 1 | VMI Raw | 8.41* | .497 | 24.66 | - | .497 |
| 2 | PRS Tot. | 5.57* | .578 | 33.40 | 8.74 | .473 |
| 3 | CARE | 2.90* | .602 | 36.18 | 2.78 | -.262 |
| 4 | DAP Raw | .89 | .610 | 37.19 | 1.01 | .330 |
| 5 | GPST | .27 | .612 | 37.42 | 0.23 | .329 |

*Denotes variables significant at $P \leq .05$.

Predicting end of grade one general academic standing by means of the complete battery of tests (i.e. including the MRT) was somewhat more successful than predicting the MRT results with 51.27% (total group), 46.09% (boys) and 56.02% (girls) of the total variance accounted for (see Table 18). The strongest predictor in all cases was the MRT. Had this test been omitted, the total variance accounted for would have been reduced by approximately 10%. It is interesting to note that the VMI contributed significantly to the regression equation for the total group and for girls whereas only the MRT and the PRS made significant contributions in the regression analysis for boys.

Classificational Analyses

Considering the trend evident in the review of the literature, this study would not be complete without looking at the data from a classificational point of view. Rather than discussing the degree of relationship between a predictor and a criterion, classificational analysis looks at the actual number of cases which would either be correctly or incorrectly identified given a specific set of criteria. For the purposes of this analysis roughly the bottom 25% of the group will be designated as "at risk". It is recognized that this is an arbitrary decision. However, given that: 1) the percentage of children expected to suffer from "learning disabilities" ranges anywhere from 5 to 15 percent (Eaves, Kendall, and Crichton, 1974) depending upon the definition employed and, 2) the concept of "at risk" used

Table 18

Multiple Regression with Grade One General Academic Standing as Criterion

Total Group

| Step | Variable | F Value | Mult.R | % Variance | %Variance Added | Simple r |
|---|---|---|---|---|---|---|
| 1 | PRS Tot. | 19.91* | .573 | 32.86 | - | .573 |
| 2 | VMI Raw | 5.13* | .637 | 40.61 | 7.75 | .512 |
| 3 | DAP Raw | .97 | .645 | 41.57 | 0.96 | .382 |
| 4 | GPST | .36 | .646 | 41.68 | 0.11 | .384 |
| 5 | MRT | 32.66* | .716 | 51.27 | 9.59 | .604 |
| | | | Boys | | | |
| 1 | PRS Tot. | 10.58* | .560 | 31.27 | - | .560 |
| 2 | VMI Raw | .79 | .592 | 35.05 | 3.78 | .456 |
| 3 | GPST | .30 | .593 | 35.22 | 0.17 | .364 |
| 4 | DAP Raw | .002 | .594 | 35.28 | 0.07 | .321 |
| 5 | MRT | 14.04* | .679 | 46.09 | 10.81 | .566 |
| | | | Girls | | | |
| 1 | PRS Tot. | 7.97* | .566 | 31.99 | - | .566 |
| 2 | VMI Raw | 5.44* | .670 | 44.89 | 12.90 | .559 |
| 3 | DAP Raw | 1.33 | .685 | 46.88 | 1.99 | .381 |
| 4 | GPST | 0.03 | .685 | 46.92 | 0.04 | .372 |
| 5 | MRT | 15.52* | .748 | 56.02 | 9.10 | .648 |

*Denotes variables significant at $P \leq .05$

in this study includes all those children who might experience learning difficulties for whatever reason (disabilities, low ability, poor motivation, emotional disturbance, lack of environmental stimulation, physical disabilities), the concept of designating the lowest quartile as the "at risk" group may not be greatly exaggerated. The cutoff scores for the predictor variables, likewise, are chosen to define, as closely as possible, the lowest 20 to 25 percent of the group. Using this criterion for "at risk", tables 19 to 23 show the percentage of children who are either correctly or incorrectly labled as a function of the prediction procedure used.

The overall hit rate for the PRS (Table 19) would seem to be quite impressive at 77.3%. Out of 172 children only 39 were incorrectly labeled. However, out of the 43 children who fell in the lowest quartile at the end of grade one, only 23 (53.5%) were identified by this procedure. Of those children who fell in the average to superior range only 19 out of 129 (14.7%) were misclassified. Another way of analysing this data is to ascertain the proportion of children correctly labeled by the procedure as either "at risk" or "not at risk". From the table it can readily be seen that of the 42 children labeled "at risk" by the PRS only 23 (54.8% had difficulty in grade one; of the 130 children considered "not at risk" 20 (15.4%) fell in the lowest quartile by the end of grade one. Stated another way, the chances of not having difficulty if labeled "at risk" are about 50-50; the chances

of having difficulty if labeled "not at risk" are about 1 in 7.

As a predictor the DAP fares even less well than the PRS with a false negative rate in excess of 60% and a false positive rate approaching 20% (Table 20). In other words, this screening instrument failed to identify almost 60% of those children who had subsequent difficulties and falsely labeled as "at risk" almost 20% of the children who, in fact, had no future problems. Using a horizontal analysis, for every 10 children labeled "at risk" only 4 would have future difficulty and for those children considered "not at risk" approximately 2 out of every 10 would have problems.

Of the three remaining screening instruments, the GPST and the VMI yield results only marginally different from the DAP (see Tables 21 & 22). Of all predictors, the MRT has the best overall hit rate (81.4%) and the smallest percentage of false negatives (44.2%) and false positives (10.1%) (See Table 23.)

Table 19

Prediction of grade one academic standing using the Pupil Rating Scale.

| PRS | Grade One Performance Low (3-8) | Average (9-11) | Superior (12-15) | Total |
|---|---|---|---|---|
| At Risk (13-34) | 23 (53.5%) | 19 (14.7%) | 0 | 42 |
| Average (35-48) | 20 (46.5%) | 67 (72.9%) | 27 | 114 |
| Superior (49-60) | 0 | 6 (12.4%) | 10 | 16 |
| | 43 | 129* | | 172 |

*Figure represents total number of average and above average pupils.

True Positives = 23/43 (53.5%)
False Negatives = 20/43 (46.4%)
Hits = 133/172 (77.3%)

False Positives = 19/129 (14.7%)
True Negatives = 110/129 (85.3%)
Misses = 39/172 (22.7%)

Table 20

Prediction of grade one academic standing using the Draw a Person Test.

| DAP | Grade One Performance Low (3-8) | Average (9-11) | Superior (12-15) | Total |
|---|---|---|---|---|
| At Risk (5-11) | 16 (37.2%) | 21 (17.8%)* | 2 | 39 |
| Average (12-23) | 25 (58.1%) | 65 (73.7%) | 30 | 120 |
| Superior (24-34) | 2 (4.7%) | 6 (8.5%) | 5 | 13 |
| | 43 | 129 | | 172 |

*The average and superior categories are grouped for the calculation of these percentages.

True Positives = 16/43 (37.2%)
False Negatives = 27/43 (62.8%)
Hits = 122/172 (70.9%)

False Positives = 23/129 (17.8%)
True Negatives = 106/129 (82.2%)
Misses = 50/72 (29.1%)

Table 21

Prediction of grade one academic standing using the Group Perceptual Screening Test.

| GPST | Grade One Performance | | | | | |
|---|---|---|---|---|---|---|
| | Low (3-8) | Average (9-11) | | Superior (12-15) | | Total |
| At Risk (3-7) | 20 (46.5%) | 16 | (16.3%)* | 5 | | 41 |
| Average (8-11) | 23 (53.5%) | 58 | (61.2%) | 21 | | 102 |
| Superior (12-14) | 0 | 18 | (22.5%) | 11 | | 29 |
| | 43 | 129 | | | | 172 |

*The average and superior categories are grouped for the calculation of these percentages.

True Positives = 20/43 (46.5%)
False Negatives = 23/43 (53.5%)
Hits = 128/172 (74.4%)

False Positives = 21/129 (16.3%)
True Negatives = 108/129 (83.7%)
Misses = 44/172 (25.6%)

Table 22

Prediction of grade one academic standing using the Test of Visual-Motor Integration.

| VMI | Grade One Performance | | | | | |
|---|---|---|---|---|---|---|
| | Low (3-8) | Average (9-11) | | Superior (12-15) | | Total |
| At Risk (4-8) | 13 (30.2%) | 12 | (10.1%)* | 1 | | 26 |
| Average (9-11) | 27 (62.8%) | 56 | (57.4%) | 18 | | 101 |
| Superior (12-16) | 3 (7.0%) | 24 | (32.5%) | 18 | | 45 |
| | 43 | 129 | | | | 172 |

*The average and superior categories are grouped for the calculation of these percentages.

True Positives = 13/43 (30.2%)
False Negatives = 30/43 (69.8%)
Hits = 129/172 (75.0%)

False Positives = 13/129 (10.1%)
True Negatives = 116/129 (89.9%)
Misses = 43/172 (25.0%)

Table 23

Prediction of grade one academic standing using the Metropolitan Readiness Test.

| MRT | Grade One Performance | | | | |
|---|---|---|---|---|---|
| | Low (3-8) | Average (9-11) | | Superior (12-15) | Total |
| At Risk (20-46) | 24 (55.8%) | 12 | (10.1%)* | 1 | 37 |
| Average (47-67) | 19 (44.2%) | 72 | (73.6%) | 23 | 114 |
| Superior (68-73) | 0 | 8 | (16.3%) | 13 | 21 |
| | 43 | 129 | | | 172 |

*The average and superior categories are grouped for the calculation of these percentages.

True Positives = 24/43 (55.8%)
False Negatives = 19/43 (44.2%)
Hits = 140/172 (81.4%)

False Positives = 13/129 (10.1%)
True Negatives = 116/129 (89.9%)
Misses = 32/172 (18.6%)

## CHAPTER 5

## Discussion and Conclusions

The purpose of this study was to examine the use of teacher evaluations, certain psychometric devices and a specific readiness test in the prediction of grade one academic achievement. There were two main focii: 1) to investigate the relationship between the various predictor and criterion variables, and 2) to ascertain the relative accuracy of prediction which would result from employing any of these procedures. A step-wise multiple regression procedure was used to determine which variables were significant predictors and which were, in fact, redundant; a classificational analysis was used to look at the question of predictive efficiency.

The results of this study revealed that the procedures under investigation were all significantly correlated with the criterion variables (MRT, Grade 1 academic standing). Furthermore, a stronger relationship was noted for girls than for boys. However, the simple correlations at no time accounted for more than 42% of the variance and, although using the predictor variables in multiple correlation did increase this figure, even the most significant result, the multiple correlation with grade 1 academic standing for for girls (R=.748), accounted for only 56% of the total variance. These results indicate that the predictor variables do not possess sufficient predictive validity to warrant their use either singly or in combination in making placement decisions about individual

children. The results of the classificational analyses support this conclusion.

The remainder of this chapter is divided into two parts: 1) a discussion of these results and their implications, and 2) recommendations and suggestions for further research arising from this study.

## Discussion and Implications

Sex differences noted by a number of researchers (Feshbach et al., 1977; Keogh & Smith, 1970; Shipp & Louden, 1964) were also evident in the present study. Girls were significantly younger than boys and, in spite of this age discrepancy, generally scored higher on all of the variables. There were significant differences in teacher evaluations (both on the simple rating scale and on the PRS), two of the paper and pencil tasks administered at the kindergarten level (GPST, DAP) and the academic standing at the end of grade one. The most significant differences were found on the DAP test and subtest 3 of the PRS (personal-social behavior). Interestingly, boys and girls did not differ significantly on the VMI.

The question as to why some children succeed and others fail and, offshoots of this question, why boys outnumber girls in learning disabilities classes and, more specifically in terms of this study, why girls seem to be more proficient in school oriented tasks have frequently been raised but, to this researcher's knowledge, no completely satisfactory explanation

has as yet been offered.

One theory prevalent in much recent research relates academic performance to the maturation of the central nervous system and more specifically to left hemispheric and cross-modal functioning. In terms of this theory, the results of the present study would seem to indicate that girls'central nervous systems develop more quickly than boys'. This conclusion seems reasonable in light of the fact that in the area of physical growth as well girls precede boys. On the other hand, neurological development may not follow the same pattern as physical development. Boys in other cultures (Japan for example) seem to develop intellectually at very early ages. Perhaps speculation as to the reasons for these noted sex differences should not be terminated too readily. As Keogh (1974) points out:

> We are uncertain as to the importance of etiology, symptom patterns, and prognosis. We have designed and implemented programs which are often confounded and confused. We rely on enthusiasm and faith instead of hard thinking and realistic evaluation....The real problem is that we don't know why or for what reasons failure or success occur. (227-228)

In keeping with this philosophy, speculation as to a possible explanation for these noted sex differences will be elucidated here and related to the discussion which will follow.

The fact that girls generally receive better scores on paper and pencil tasks especially at an early age may be related to a practice effect. In our society, girls are far more likely than boys to be using crayons, scissors, glue and the like from a very early age. For many boys, kindergarten may be their

first exposure to such demands. In addition to pure awkwardness with the "tools of the trade", so to speak, a lack of motivation is also often evident. It would seem, from this researcher's observations, that boys might rather produce three dimensional models with blocks, lego, or tinkertoys than symbolic representations in two dimensional space. The DAP, for example, may have been executed in a perfunctory manner by the boys as it held little interest for them whereas the girls may have taken a great deal more time to produce an aesthetically pleasing result. In other words, as Keogh (1971) cautioned, success may indicate an ability but failure does not necessarily indicate lack of ability. The fact that on the VMI, a more highly structured drawing task, performance was essentially the same for boys and girls may partially support this thesis. Perhaps if the task had been to produce some sort of structure out of lego, the boys would have scored significantly higher than the girls.

The foregoing discussion in no way negates or denies the role of left hemispheric and cross-modal functioning in such tasks as reading, writing, and spelling. It is simply advanced as a possible explanation as to why girls, even at a younger age, seemingly find these tasks easier than boys. The differential development of the central nervous system may not so much be a product of time as of practice. The fact that there was no significant relationship between age and end of grade one

academic standing either for boys or for girls and that differences between boys and girls seem to attenuate over time lends credence to this theory.

The paper and pencil tasks administered to the kindergarten children then are likely biased in favor of the girls. Indeed, due to the hypothesized "practice effect" girls' scores may be inflated to a similar degree that boys' scores are deflated. However, in that such tasks reflect a basic orientation of the school system in general, they are likely to be quite predictive of future performance. A glance at the correlation matrix for the total group (Table 14) indeed reveals significant correlations with grade one academic standing for all such variables. The correlations, ranging from .37 to .50 are similar to those reported by other researchers (Duffey, Ritter, & Fedner, 1976; Feshbach, Adelman, & Fuller, 1974; Flynn & Flynn, 1978). Calculating the correlations separately by sex, however, reveals somewhat stronger relatinships between predictor and criterion for girls than for boys (r range for boys = .32 to .43, for girls the range is from .37 to .56). In keeping with the theory expressed previously, paper and pencil tasks may be a more accurate (perhaps somewhat inflated) representation of girls' basic visual-motor integration skills in that they are familiar with such activities and generally quite motivated to carry them out. The same may not be true for boys whose initial performance is possibly negatively influenced by general lack of

interest in and experience with such activities. Furthermore, exposure to these tasks over the kindergarten and grade one year likely results in rapid changes in skill level for boys whereas girls' skills in these areas likely remain fairly stable over this time period owing to the fact that they were quite highly developed to begin with.

The second instrument on which boys and girls differed substantially was subtest 3 of the PRS. There were no significant differences on subtest 1 (auditory comprehension) and the differences on subtest 2 (spoken language) did not achieve significance at the .05 level. In other words, there did not seem to be any differences between boys and girls in their comprehension of language or in their ability to follow oral directions. On the other hand, girls seemed to be somewhat more able to express themselves verbally. These subtests were equally strong predictors for both groups in terms of end of grade one performance (Tables 15 & 16). It was on subtest 3 that the most significant difference in means occurred (P=.001). Teachers perceived girls to be significantly more settled, to display more school appropriate behaviors and hence, as evident from the scores on the simple rating scale as well, to be significantly more ready to enter grade one than boys. Interestingly, the correlations between subtest 3 of the PRS and both the MRT and the end of grade one academic standing were much stronger for girls (.49 and .50 respectively) than for boys (.28 in both

cases). In other words, teachers' perceptions about girls' social-emotional development were more predictive than their perceptions about boys. These discrepant correlations might be due to bias on the part of the teachers or to a weaker relationship between boys' behavior and future performance. In this researcher's opinion there is no justification in assuming that teachers, who are highly trained, sensitive and professional individuals, are either unable or unwilling to rate boys as accurately as girls. The explanation, then, must lie in the second hypothesis, a weaker link between behavior and future performance for boys. As any preschool teacher knows, girls and boys often exhibit totally different sets of behaviors when they first enter the kindergarten classroom. On the average, girls generally start out with many well established school-appropriate skills and behaviors and their progress continues in a consistent and linear fashion. Many boys, on the other hand, come to kindergarten with untried skills and inappropriate school behaviors. It is this group of children who are probably most affected by the demands of the new situation and undergo the most significant changes, making predictions difficult and generally less accurate. As with physical development, there are periods of rapid growth interspersed with plateaus. The apparent weaker link between behavior and performance for boys is likely simply a reflection of that tremendous potential for change so evident in young

children. In other words, as Flynn and Flynn (1978) noted, the problems related to prediction likely inhere more in the nature of the child than in the instruments used.

For all practical purposes, using the predictor variables in multiple correlation did not significantly change the prediction picture when the MRT was the criterion variable. Only two variables, the VMI and the PRS, contributed significantly to the multiple correlation when the total group was considered. When the groups were analysed separately by sex a third variable, CARE, contributed at a significant level. This variable was included to test the hypothesis that children in daycare would perform differently from children in homecare. Interestingly, boys and girls seemed to respond differentially to this condition. The correlation between CARE and MRT results for boys was .20 ($P = .027$) and for girls it was -.25 ($P = .014$). From these correlations it would seem that daycare (i.e. the higher score on the CARE variable) was associated with higher scores on the MRT for boys; whereas for girls, homecare (i.e. the lower score on the CARE variable) was related to higher scores on the MRT. As the number of children in daycare situations was small (26 boys and 17 girls), interpretations must be made cautiously, but these discrepant correlations are not inconsistent with the theory previously elaborated. In other words, daycare, in that it is an institutional setting, imposes its structure and limitations on the boys before they begin kindergarten thus

giving these children a headstart, so to speak, in the process of adaptation to the school learning situation. Girls, on the other hand, seem to make better progress if they remain in their home environment prior to entering kindergarten. Perhaps they receive more structure and instruction at home under the tutelage of their mothers. Being constantly in the presence of their most immediate sex role model possibly ensures an ongoing and intense learning situation which is denied boys.

Using the kindergarten variables plus the MRT in a multiple correlational analysis with grade one academic standing as the criterion increased the explained variance substantially from roughly 33% to 52% (total group). Although, as expected, a higher correlation was achieved for girls than for boys, there was essentially no difference in the order in which the variables entered the equation. In all cases (total group, boys, girls) the MRT was the strongest predictor followed by the PRS. When the group was considered as a whole the VMI made a significant contribution to the overall correlation as it did in the case of the girls. From these results it would appear that each type of predictor variable under analysis in this study (teacher evaluations, psychometric procedures and readiness tests) contributed relevant information to the total prediction equation. Although in the case of the boys only the MRT and the PRS were significant predictors at the .05 level, the VMI was the third variable to enter the equation.

Although it was evident from the correlational analyses that using these predictor variables would not yield highly accurate predictions, classificational analyses were undertaken in this research study to clarify where prediction errors would occur given a specific set of criteria. The results indicated that, of all the predictor variables, the PRS and the MRT identified the greatest percentage of children (53.5% and 55.8% respectively) who had subsequent difficulties and made the fewest prediction errors. The accuracy of prediction, however, was far too low to justify using the results to make placement decisions about individual children. If labeled "at risk" by either of these instruments, there was still almost a 50-50 chance that the child would achieve in the average range by the end of grade one. In addition, roughly 45% of those children who had subsequent difficulties were not identified by these instruments. Although these results were somewhat stronger than those achieved by Feshbach et al. (1974) using both the SRS, a teacher rating scale, and a modified version of the Predictive Index (See Tables 1, 2, & 7), they were much less impressive than those of a number of other researchers including Jansky and de Hirsch (1972) and Satz et al. (1976) who used a battery approach to prediction (Tables 6 & 9) and Lessler and Bridges (1973) who used single readiness tests as predictors (Tables 3, 4, & 5).

Recommendations and Suggestions for Future Research

The results of the present research study support the findings of those earlier investigations which found significant relationships between task performance in kindergarten and later academic achievement. However, the regression and classificational analyses which were conducted bring into question the utility of such information. It is evident that, in spite of significant correlations between predictor and criterion variables, there would be a large margin for error if any of the instruments under investigation was used to make actual placement decisions about individual kindergarten children. In this study the fewest prediction errors were made for those children who scored in the superior range on the early predictors. Almost all of these children scored in the average or superior range at the end of grade one. In other words, if a child is doing very well in kindergarten the possibility of falling into the "at risk" group in grade one is very slight. Similarly, children who performed at the lowest level in kindergarten were not likely to be superior students by the end of grade one, however, there was a strong possibility that they would fall in the average group. With those children predicted to be "average" the greatest proportion (about .66) indeed fell in the average range a year later. The remaining third, however, was about equally divided between the "at risk" and the "superior" groups.

Perhaps the time has come to reevaluate our need to predict the future. Although it might be comforting to know that using

a specific battery of tests would enable us to state with a high degree of accuracy that a given child would or would not have a successful grade one experience, this would be to deny that tremendous human potential for change and development. Predicting an outcome presupposes the ability to isolate and measure all variables impinging upon that situation as well as make accurate and reliable statements about the laws governing the complete process. In the case of human growth and development not only are the variables so numerous and diverse as to be virtually indescribable, the child comes to the situation with a unique personality and set of behavioral responses. Although the growth process (cognitive as well as physical) is predictable in that it seems to proceed by stages in a sequential manner, it is unpredictable in that different children progress at different rates through these stages. Children between the ages of five and seven are in a state of tremendous flux. Inter- and intra-variability is the norm for this age group (Flynn & Flynn, 1978; Snyder & Pope, 1972) making accurate prediction as to future level of functioning very difficult.

As mentioned previously, using norm-referenced tests as a basis for predicting future academic achievement might be an invalid procedure, especially when dealing with young children, for two reasons: 1) variable motivation and 2) extremely variable developmental rate among this age group. In addition, there is something inherently negativistic and discouraging about any

testing procedure which ensures that 50% of the group is "below average". Rather than looking for deficits in children, Keogh (1971) suggests we should be looking for competencies. In a similar vein, Wendt (1978) argues that criterion- rather than norm-referenced tests should be used in assessing young children. Knowing and stating what a child is able to do is far more positive than enumerating suspected deficits or lags. In addition, this approach allows for realistic program planning. Activities can be structured in such a way as to take advantage of a child's strengths in order to facilitate growth in weaker areas. As Jansky (1978) states:

> What may determine the child's ultimate success or failure in reading are the subtle interactions of areas of immaturity and compensatory strengths wherever they occur. (389)

Another direction which future research might take would be to design research projects to thoroughly investigate the performance and prediction differences between boys and girls. We need to know why girls seem to adjust more readily to the learning situation. Is it a biological phenomenon or is it, as was hypothesized by this researcher, culturally determined? Would it make a difference if the kindergarten teacher was a man instead of a woman? Are boys really "immature" or are our expectations unsuitable? The answers to these questions would facilitate policy decisions and program development.

A third area which this research suggests would be interesting to investigate is the effect of daycare on school perfor-

mance. The results of the present study suggest that boys and girls may respond differentially to this condition. As early as 1965 Ilg, Ames, and Appel suggested that the greatest source of school problems was grade overplacement and that girls should be fully 6 on entering grade one and boys should be a full six months older. In a later paper (Ilg, Ames, Haines, & Gillespie, 1978) this was modified to allow for a two year waiting period before beginning grade one if a child should be deemed to be developmentally dagging to that degree. If, however, future research supports the finding that institutional settings are beneficial for boys, radical changes in our thinking would be indicated. Rather than keeping a boy at home for a longer period of time and waiting for him to "mature", it might be more efficacious for him to start school at an even earlier age and allow for a one or two year pre-kindergarten developmental period.

# BIBLIOGRAPHY


Adelman, H. S., & Feshbach, S. Predicting reading failure: Beyond the readiness model. Exceptional Children, 1970, 37, 349-354.

Ames, S. G., Becker, L. D., & Dalton, S. The predictive validity of the Stanford Early School Achievement Test. Educational and Psychological Measurement, 1977, 39, 505-507.

Ashmore, R. J., & Snyder, R. T. Relationship of visual and auditory short-term memory to later reading achievement. Perceptual and Motor Skills, 1980, 51, 15-18.

Beery, K. E. Developmental Test of Visual-Motor Integration: Administration and scoring manual. Chicago: Follett Publishing Co., 1967.

Benton, A. L. Developmental dyslexia: Neurological aspects. In: Advances in Neurology. Edited by W. J. Friedlander. New York: Raven Press, 1975, 1-47.

Berk, R. A. What's wrong with using grade-equivalent scores to identify LD children? Academic Therapy, 1981, 17, 133-140.

Book, R. M. Predicting reading failure: A screening battery for kindergarten children. Journal of Learning Disabilities, 1974, 7, 43-47.

Book, R. M. Identification of educationally at-risk children during the kindergarten year: A four-year follow-up. Psychology in the Schools, 1980, 17, 153-158.

Bremer, N. D. Do readiness tests predict success in reading? Elementary School Journal, 1959, 59, 222-224.

Byran, T. S., & McGrady, H. J. Use of a teacher rating scale. Journal of Learning Disabilities, 1972, 5, 199-206.

Chalafant, J. C., & Scheffelin, M. A. Central processing dysfunctions in children: Review of research. Washington, D. C.: U. S. Government Printing Office, 1969.

Chissom, B. A. A factor-analytic study of the relationship of motor factors to academic criteria for first- and third-grade boys. Child Development, 1971, 42, 1133-1143.

Colarusso, R., Gill, S., Plankenhorn, A., & Brooks, R. Predicting first-grade achievement through formal testing of 5-year-old high-risk children. Journal of Special Education, 1980, 14, 355-363.

Colligan, R. C., Concurrent validity of the Myklebust Pupil Rating Scale in a kindergarten population. Journal of Learning Disabilities, 1977, 10, 317-320.

Colligan, R. Predictive utility of the Myklebust Pupil Rating Scale: A two-year follow-up. Journal of Learning Disabilities, 1979, 12, 264-267.

deHirsch, K., Jansky, J. J., & Langford, W. S. Predicting reading failure. New York: Harper and Row, 1966.

Duffey, J. B., Ritter, D. R., & Fedner, M. Developmental Test of Visual-Motor Integration and the Goodenough Draw-a-Man Test as predictors of academic success. Perceptual and Motor Skills, 1976, 43, 543-546.

Dunleavy, R. A., Hansen, J. L., Szasz, C. W., & Baade, L. E. Early kindergarten identification of academically not-ready children by use of human figure drawing developmental score. Psychology in the Schools, 1981, 18, 35-38.

Eaves, L. C., Kendall, D. C., & Crichton, J. U. The early detection of minimal brain dysfunction. Journal of Learning Disabilities, 1972, 5, 454-462.

Eaves, L. C., Kendall, D. C., & Crichton, J. U. The early identification of learning disabilities: A follow-up study. Journal of Learning Disabilities, 1974, 7, 632-638.

Eisenberg, L. The epidemiology of reading retardation and a program for preventive intervention. In: The disabled reader: Education of the dyslexic child. Edited by J. Money. Baltimore: Johns Hopkins Press, 1966, 3-20.

Eisenberg, L. Definitions of dyslexia: Their consequences for research and policy. In: Dyslexia: An appraisal of current knowledge. Edited by A. L. Benton and D. Pearl. New York: Oxford University Press, 1978, 31-42.

Federici, L., Sims, H., & Bashian, A. Use of the Meeting Street School Screening Test and the Myklebust Pupil Rating Scale with first-grade black urban children. Psychology in the Schools, 1976, 13, 386-389.

Ferinden, W. E., Jacobsen, S., & Linden, N. J. Early identification of learning disabilities. Journal of Learning Disabilities, 1970, 3, 589-593.

Feshbach, S., Adelman, H., & Fuller, W. W. Early identification of children with high risk of reading failure. Journal of Learning Disabilities, 1974, 7, 639-644.

Feshbach, S., Adelman, H., & Fuller, W. W. Prediction of reading and related academic problems. Journal of Educational Psychology, 1977, 69, 229-308.

Fisk, R. Predicting group membership for learning disabled and control subjects with selected cognitive and affective measures. Masters thesis, University of Alberta, 1979.

Fletcher, J. M. & Satz, P. Unitary deficit hypotheses of reading disability: Has Vellutino led us astray? Journal of Learning Disabilities, 1979, 12, 155-159.

Flynn, T. M., & Flynn, L. A. Evaluation of the predictive ability of five screening measures administered during kindergarten. Journal of Experimental Education, 1978, 46, 65-70.

Friedman, R., Fuerth, J. H., & Forsythe, A. B. A brief screening battery for predicting school achievement at ages seven and nine years. Psychology in the Schools, 1980, 17, 340-345.

Gaddes, W. Learning disabilities: Prevalence estimates and the need for definition. In: The neuropsychology of learning disorders. Edited by R. Knights and D. J. Bakker. Baltimore: University Park Press, 1976, 3-24.

Gallerani, D., O'Regan, M., & Reinherz, H. Prekindergarten screening: How well does it predict readiness for first grade? Psychology in the Schools, 1982, 19, 175-182.

Georgas, J. G., & Papadopoulou, E. The Harris-Goodenough and the Developmental Form Sequence with five-year-old Greek children. Perceptual and Motor Skills, 1968, 26, 352-354.

Gredler, G. R. A look at some important factors in assessing readiness for school. Journal of Learning Disabilities, 1978, 11, 284-290.

Gruen, R. S. Prediction of end-of-year reading achievement for first and third grade pupils. Proceedings of the American Psychological Association, 1972, 563-64.

Harris, D. B. Children's drawings as measures of intellectual maturity. New York: Harcourt, Brace and World, 1963.

Harris, D. B. The Goodenough-Harris drawing test: Manual. New York: Harcourt, Brace and World, 1963.

Harris, L. J. Sex differences in spatial ability: Possible environmental, genetic and neurological factors. In: Asymmetrical function of the brain. Edited by M. Kinsbourne. Cambridge: Cambridge University Press, 1978.

Hase, H. D. Predicting performance in the first grade with the First-Grade Screening Test. Psychology in the Schools, 1977, 14, 407-412.

Hynd, G. W. & Obrzut, J. E. Reconceptualizing cerebral dominance: Implications for reading and learning disabled children. Journal of Special Education, 1981, 15, 447-457.

Ilg, F. L., & Ames, L. B. School readiness. New York: Harper and Row, 1965.

Ilg, F., Ames, L. B., & Appel, R. J. School readiness as evaluated by Gessel development, visual and predictive tests. Genetic Psychology Monographs, 1965, 71, 61-91.

Ilg, F. L., Ames, L. B., Haines, J., & Gillespie, C. School Readiness. New York: Harper and Row, 1978.

Jansky, J. J. A critical review of some developmental and predictive precursors of reading disabilities. In: Dyslexia: An appraisal of current knowledge. Edited by A. L. Benton and D. Pearl. New York: Oxford University Press, 1978, 379-394.

Jansky, J. and deHirsch, K. Preventing reading failure: Prediction, diagnosis, intervention. New York: Harper and Row, 1972.

Kelly, G. R. Group perceptual screening at first grade level. Journal of Learning Disabilities, 1970, 13, 640-644.

Keogh, B. K. A compensatory model for psycho-educational evaluation of children with learning disorders. Journal of Learning Disabilities, 1971, 4, 544-548.

Keogh, B. K. Optometric vision training programs for children with learning disabilities. Journal of Learning Disabilities, 1974, 7, 219-231.

Keogh, B., & Becker, L. Early detection of learning problems: Cautions and guidelines. Exceptional Children, 1973, 40, 5-11.

Keogh, B. K., & Smith, C. E. Early identification of educationally high potential and high risk children. Journal of School Psychology, 1970, 8, 285-290.

Keogh, B. K., Tchir, C., & Windeguth-Behn, A. Teachers' perceptions of educationally high risk children. Journal of Learning Disabilities, 1974, 7, 367-374.

Kermonian, S. B. Teacher appraisal of first grade readiness. Elementary English, 1962, 39, 196-201.

Kinsbourne, M., & Hiscock, M. Cerebral lateralization and cognitive development. Education and the brain. Chicago: University of Chicago, 1978.

Klein, A. The validity of the Screening Test of Academic Readiness in predicting achievement in first and second grades. Educational and Psychological Measurement, 1977, 37, 493-499.

Knights, R., M., & Bakker, D. J. (Eds.) The neuropsychology of learning disorders: Theoretical approaches. Baltimore: University Park Press, 1976.

Larsen, S. C., Rogers, D., & Sowell, V. The use of selected perceptual tests in differentiating between normal and learning disabled children. Journal of Learning Disabilities, 1976, 9, 85-90.

Lehman, J., & Breen, M. J. A comparative analysis of the Bender-Gestalt and Beery/Buktenica Test of Visual-Motor Integration as a function of grade level for regular education students. Psychology in the Schools, 1982, 19, 52-54.

Lessler, K., & Bridges, J. S. The prediction of learning problems in a rural setting: Can we improve on readiness tests? Journal of Learning Disabilities, 1973, 6, 90-94.

Lindquist, G. T. Preschool screening as a means of predicting later reading achievement. Journal of Learning Disabilities, 1982, 15, 331-332.

Lindsay, G. A., & Wedell, K. The early identification of educationally 'at risk' children revisited. Journal of Learning Disabilities, 1982, 15, 212-217.

Lipson, A. M. Catching them early. Academic Therapy, 1981, 16, 457-463,

Maitland, S., Nadeau, J. B. E., & Nadeau, G. Early school screening practices. Journal of Learning Disabilities, 1974, 10, 645-649.

Margolis, H., Sheridan, R., & Lemanowicz, J. The efficiency of Myklebust's Pupil Rating Scale for detecting reading and arithmetic difficulties. Journal of Learning Disabilities, 1981, 14, 267-268.

Mattis, S. Dyslexia syndromes: A working hypothesis that works. In: Dyslexia: An appraisal of current knowledge. Edited by A. L. Benton and D. Pearl. New York: Oxford University Press, 1978, 45-58.

Meehl, P., & Rosen, A. Antecedent probability and the efficiency of psychometric signs, patterns or cutting scores. Psychological Bulletin, 1955, 52, 194-216.

Mercer, C. D., Algozzine, B., & Trifiletti, J. J. Early identification: Issues and considerations. Exceptional Children, 1979, 46, 52-54.

Money, J. On learning and not learning to read. In: The disabled reader: Education of the dyslexic child. Edited by J. Money. Baltimore; Johns Hopkins Press, 1966, 21-40.

Myklebust, H. The Pupil Rating Scale: Screening for learning disabilities. New York: Grune and Stratton, 1971.

Myklebust, H. R. What is the future for LD? Journal of Learning Disabilities, 1980, 13, 468-471.

Newell, D., & Rugel, R. P. Hemispheric specialization in normal and disabled readers. Journal of Learning Disabilities, 1981, 14, 296-297.

Nichta, L. J., Federici, L., & Schuerger, J. The Screening Test of Academic Readiness (STAR) as a predictor of third-grade achievement. Psychology in the Schools, 1982, 19, 190-193.

Pihl, R. O., & Nagy, K. A. The applicability of the Myklebust Pupil Rating Scale. Journal of Learning Disabilities, 1980, 13, 109-113.

Porter, G. L., & Binder, D. M. A pilot study of visual-motor development inter-test reliability: The Beery Developmental Test of Visual-Motor Integration and the Bender Visual-Motor Gestalt Test. Journal of Learning Disabilities, 1981, 14, 124-127.

Proger, B. B. The Pupil Rating Scale: Screening for learning disabilities. Journal of Special Education, 1973, 7, 311-317.

Pryzwansky, W. B. The use of the Developmental Test of Visual-Motor Integration as a group screening instrument. Psychology in the Schools, 1977, 14, 419-422.

Raccioppi, R. M. The promise of Public Law 94-142: Issues in the diagnostic process. Journal of Learning Disabilities, 1982, 15, 218-220.

Reinherz, H., & Griffin, C. L. Identifying children at risk: A first step to prevention. Health Education, 1977, 8, 14-16.

Reisman, J., & Yomokoski, T. Can intelligence be estimated from drawings of a man? Journal of School Psychology, 1973, 11, 239-244.

Rewilak, D., & Janzen, H. L. A discriminant function analysis of cognitive and non-cognitive measures for learning disabled and normally-achieving students. Manuscript, University of Alberta, Department of Educational Psychology, 1982.

Reynolds, C. R. An examination for bias in a preschool test battery across race and sex. Journal of Educational Measurement, 1980, 17, 137-146.

Reynolds, C. R. Differential construct validity of a preschool battery for blacks, whites, males, and females. Journal of School Psychology, 1980, 18, 112-125.

Reynolds, C. R., Wright, D,. & Wilkinson, W. A. Incremental validity of the Test for Auditory Comprehension of Language and the Developmental Test of Visual-Motor Integration. Educational and Psychological Measurement, 1980, 40, 503-507.

Richardson, C. M. Learning disability procedures: A human rights perspective. Journal of Learning Disabilities, 1981, 14, 7-8.

Ritter, D., Duffey, J. B., & Fischman, R. Comparisons of the intellectual estimates of the Draw-a-Man Test and the Stanford-Binet for kindergarten children. Psychology in the Schools, 1974, 11, 412-415.

Rogolsky, M. M. Screening kindergarten children: A review and recommendations. Journal of School Psychology, 1968, 7, 18-27.

Rourke, B. P. Reading retardation in children: Developmental lag or deficit? In: Neuropsychology of learning disorders: Theoretical approaches. Edited by R. M. Knights and D. J. Bakker. Baltimore: University Park Press, 1976, 125-137.

Ryckman, D. B., & Rentfrow, R. The Beery Developmental Test of Visual-Motor Integration: An investigation of reliability. Journal of Learning Disabilities, 1971, 4, 333-334.

Sabatino, D. A., & Hayden, D. L. Variation on information processing behaviors. Journal of Learning Disabilities, 1970, 3, 404-412.

Satz, P., & Fletcher, J. M. Early screening tests: Some uses and abuses. Journal of Learning Disabilities, 1979, 12, 56-60.

Satz, P., & Fletcher, J. Minimal brain dysfunctions: An appraisal of research concepts and methods. In: Handbook of minimal brain dysfunctions: A critical view. Edited by H. Rie and E. Rie. New York: Wiley Interscience, 1980, 669-714.

Satz, P., & Friel, J. Some predictive antecedents of specific learning disability: A preliminary one-year follow-up. In: The disabled learner. Edited by P. Satz and J. Ross. Rotterdam: University Press, 1972.

Satz, P., & Friel, J. Some predictive antecedents of specific reading disability: A preliminary two-year follow-up. Journal of Learning Disabilities, 1974, 7, 437-444.

Satz, P., & Friel, J. Predictive validity of an abbreviated screening battery. Journal of Learning Disabilities, 1978, 11, 347-351.

Satz, P., Friel, J., & Rudegeair, F. Some predictive antecedents of specific reading disability: A two-, three-, and four-year follow-up. In: Aspects of Reading Acquisition. Edited by J. T. Guthrie. Baltimore: Johns Hopkins University Press, 1976, 111-140.

Satz, P., Rardin, D., & Ross, J. An evaluation of a theory of specific developmental dyslexia. Child Development, 1971, 42, 2009-2021.

Satz, P., & Sparrow, S. S. Specific developmental dyslexia: A theoretical formulation. In: Specific reading disability: Advances in theory and method. Edited by D. J. Bakker and P. Satz. Rotterdam: University Press, 1970, 17-40.

Satz, P., Taylor, H. G., Friel, J., & Fletcher, J. M. Some predictive and developmental precursors of reading disability: A six-year follow-up. In: Dyslexia: An appraisal of current knowledge. Edited by D. Pearl and A. Benton. Oxford: University Press, 1979, 313-318.

Satz, P., & van Nostrand, G. K. Developmental dyslexia: An evaluation of a theory. In: The disabled learner. Edited by P. Satz and J. Ross. Rotterdam: University Press, 1972, 121-148.

Shipp, D. E., & Loudon, M. L. The Draw-a-Man Test and achievement in the first grade. Journal of Educational Research, 1964, 57, 518-519.

Silver, A. A. Prevention. In: Dyslexia: An appraisal of current knowledge. Edited by A. L. Benton and D. Pearl. New York: Oxford University Press, 1978, 351-376.

Simner, M. L. Printing errors in kindergarten and the prediction of academic performance. Journal of Learning Disabilities, 1982, 15, 155-159.

Snyder, R., & Pope, P. Auditory and visual inadequacies in maturation at the first grade. Journal of Learning Disabilities, 1972, 5, 620-635.

Snyder, P., Snyder, R. T., & Massong, S. F. The Visual-Motor Integration Test: High interjudge reliability, high potential for diagnostic error. Psychology in the Schools, 1981, 18, 55-59.

Sparrow, S. S. & Satz, P. Dyslexia, laterality and neuropsychological development. In: Specific reading disability: Advances in theory and method. Edited by D. J. Bakker and P. Satz. Rotterdam: University Press, 1970, 17-40.

Stevenson, H. W., Parker, T., Wilkinson, A., Hegion, A., & Fish, E. Longitudinal study of individual differences in cognitive development and scholastic achievement. Journal of Educational Psychology, 1976, 68, 377-399.

Stevenson, H. W., Parker, T., Wilkinson, A., Hegion, A., & Fish, E. Predictive value of teachers' ratings of young children. Journal of Educational Psychology, 1976, 68, 507-517.

Telegdy, G. A factor analysis of four school readiness tests. Psychology in the Schools, 1974, 11, 127-133.

Telegdy, G. The effect of four readiness tests as predictors of first grade academic achievement. Psychology in the Schools, 1975, 12, 4-11.

Vellutino, F. R. The validity of perceptual deficit explanations of reading disability: A reply to Fletcher and Satz. Journal of Learning Disabilities, 1979, 12, 160-167.

Vellutino, F. R., Steger, B. M., Moyer, S. C., Harding, C. J., & Niles, J. A. Has the perceptual deficit hypothesis led us astray? Journal of Learning Disabilities, 1977, 10, 375-385.

Wedell, K. Learning and perceptuo-motor disabilities in children. London: John Wiley & Sons, 1973.

Wendt, R. N. Kindergarten entrance assessment: Is it worth the effort? Psychology in the Schools, 1978, 15, 56-62.

White, M., Batini, P., Satz, P., & Friel, J. Predictive validity of a screening battery for children 'at risk' for reading failure. British Journal of Educational Psychology, 1979, 49, 132-137.

# APPENDIX A

## Correspondence with the Schools

February 18, 1982.

To: Sherwood Park Elementary School Principals-

Dear

I have been with the County of Strathcona for the past three years as a guidance counsellor at Campbelltown school and am presently attempting to complete my degree in school psychology. My thesis topic has recently been approved by the Department of Educational Psychology at the U. of A. and has received tacit approval of the school board, subject to acceptance by individual principals.

In this regard I am forwarding you a copy of my proposal and requesting your assistance. I need eight to ten classroom groupings of kindergarten children enrolled in the English language program. The study has been designed so that classroom routine would only be minimally affected. Testing would take place late in April or early in May. Should you have any further questions please feel to contact me at 467-5143.

Should you be interested in having your E.C.S. students involved in this study, please complete the form below and return it to me as soom as possible. Your support in this matter is greatly appreciated.

Yours truly,

---

School_____

Total number of E.C.S. children enrollen in the English Language program _____

_____(Principal)

March 1, 1983.

To:

From: Mrs. Betty Nadon, Counsellor, Campbelltown School.

Re: Kindergarten-Grade 1 Prediction Study

At last this study is nearing completion. I am now requesting the classroom teacher to rate each child in terms of his general performance in the areas of: reading, written and oral language, computation and social-emotional behavior.

Detailed directions for the completion of this task are enclosed and should be studied before any attempt is made to complete the lists. Although it is not expected that this should be an arduous task, it would likely be appreciated by the teachers and would possibly increase the reliability of the results if your counsellor could assist in the completion of the forms.

Thank you again for your continued interest in and support for this project.

Kindergarten-Grade 1 Prediction Study: Guidelines for the completion of the March Performance Ratings

1. In each of the subject areas decide whether the child would best be described as "average", "below average" or "above average" and assign a number as indicated below.

2. Reconsider those children to whom you assigned a rating of 2 (below average) or 4 (above average) and decide whether any of them is decidedly inferior or superior to the rest of the children in either of those groups. If warranted place an asterisk (*) in the column next to the original rating. NOTE: Except in rare instances no more than 2 or 3 children out of a group of 25 would be expected to fall into either of these two extreme categories.

Ratings

Below Average..2
Average.....3
Above Average..4

Clarification: A rating of 2 followed by an asterisk ( 2 * ) would designate the lowest children in the class whereas a rating of 4 followed by an asterisk ( 4 *) would identify the superior group.

# APPENDIX B

## Correspondence with Parents

(Appropriate School Letterhead)

Dear Parents:

The school board has granted Mrs. Betty Nadon, counsellor at Campbelltown School, permission to conduct a research project involving kindergarten children. Our school plus three other Sherwood Park schools have agreed to participate.

The study requires that the children carry out three separate drawing tasks. They will be done in the classroom on a small group basis under the supervision of the teacher. The children will not be identified individually in the study. Only group information is required.

The purpose of the study is to attempt to verify if and to what degree age factors, visual-motor skills and/or teacher observations are related to overall academic readiness. The results obtained from the kindergarten sampling will be compared with first grade core testing results in the fall. The study should be completed by Christmas, 1982. A copy will be available at the school should you be interested in the results.

If you have any concerns or questions regarding this project please feel free to contact Mrs. Nadon at 467-5143.

Yours truly,

_____

Classroom teacher

_____

Principal

# APPENDIX C

## Comparison of Group Means: Moved versus Nonmoved

## Appendix C

### Differences in Means (Moved versus Nonmoved)

| Variable | Moved | Nonmoved | F | P Level |
|---|---|---|---|---|
| Age | 69.95 | 70.23 | .093 | .760 |
| GPST | 9.23 | 9,05 | 1.629 | .203 |
| DAP Raw Score | 17.59 | 15.66 | 2.419 | .121 |
| DAP IQ | 96.95 | 93.35 | 1.474 | .226 |
| VMI Raw Score | 10.36 | 10.18 | .172 | .679 |
| VMI Age | 68.95 | 68.10 | .228 | .634 |
| VMI Discrepancy | 24.00 | 22.91 | .357 | .551 |
| PRS Total | 37.59 | 37.81 | .016 | .901 |
| PRS Sub. 1 Aud. Comp. | 12.23 | 12.73 | .532 | .466 |
| PRS Sub. 2 Spoken Lang. | 12.32 | 12.70 | .369 | .544 |
| PRS Sub. 3 Pers.-Social | 13.05 | 12.37 | 1.213 | .272 |
| Rating | 2.09 | 2.14 | .130 | .710 |